Annual Progress Report

Meta-Analysis and Benefit Transfer at Different Levels of Aggregation: Comparing Group-Averaged and Individual-Level Models Using Hierarchical Bayesian Methods

EPA award # RD-83346001-0

Synthesis of Objectives

The aim of this project is to compare Benefit Transfer results for resource valuation flowing from Meta-Regression Models with different levels of data aggregation. We hypothesize that conventional Aggregate Level Meta-Regression Models (ALMRMs) ignore underlying individual heterogeneity in preferences and behavior and thus understate the variability in expected benefits flowing from the transfer function. Avoiding this shortfall requires working with Individual Level Meta-Regression Models (ILMRMs) based on original data. Naturally, this is costly in terms of time and – potentially – monetary resources. Thus, we expect this project to shed light on the relative merits of "going this extra mile" when amassing a dataset for meta-modeling and BT. The additional effort of contacting original authors and recouping original raw data would certainly be justifiable if the ALMRM leads to gross estimation errors compared to the ILMRM. However, if our research establishes that the ALMRM generally performs adequately to address the principal policy questions at hand, a traditional BT via aggregate MRM may suffice.

An added benefit of access to original data is the avoidance of data gaps in the ALMRM, since information on regressors not provided in printed outlets (which form the basis for an ALMRM) can be substituted using the original data. We label the resulting improved MRM the Enriched Aggregate Meta-Regression Model or EAMRM. In a nutshell, the main objective of the project is to compare the performance of the ALMRM, EAMRM, and ILMRM, using Bayesian hierarchical modeling techniques for maximum flexibility in data combination and generation of the BT function.

We originally planned to compare these different MRMs for two areas of resource valuation for which BT has been used extensively in the past: hiking and recreational fishing[1]. For each area, we planned to identify a set of suitable studies, and to contact the original authors to obtain the underlying raw data. We then planned to estimate a generic ALMRM, fill data gaps by resorting to the actual data, estimate the resulting EAMRM, and compare both models to a full-fledged ILMRM. Parallel to data collection efforts, we planned to estimate a variety of MRMs using simulated data sets.

We pursued the collection of hiking data as planned. The details of this effort are given in the next section. However, we decided to change the second application from sportsfishing to *land preservation* for a variety of reasons: (i) Land preservation is a public resource topic of growing interest to a variety of agencies, (ii) Co-PI Johnston is intimately familiar with this topic due to his primary research efforts in this field in recent years, (iii) It is a topic that has not yet been fully examined with respect to the potential of BT to inform policy decisions, and (iv) The homogeneous nature of the resulting data facilitates the estimation and comparison of our envisioned meta-models.

Progress Summary / Accomplishments

<u>Data Collection for Hiking Application</u>

Co-PI Rosenberger was in charge of collecting the hiking data. As envisioned in the project proposal, he focused on studies that used surveys to estimate the "all-or-nothing" welfare effects of site access. A total of 39 initial studies were identified by searching The Recreation Use Values Database hosted by Dr. Rosenberger at Oregon State University. The list was further adjusted to only include

[1] However, as stated in footnote 4 of the original proposal, depending on data availability and quality we reserved the right to switch to other applications if necessary.

target studies that comprised individual travel cost data, resulting in a refined set of 17 candidate studies. Official letters requesting raw data were sent to the authors of these source studies. Raw data could be obtained for seven of these studies, including four data sets covering fire effects on hiking and mountain biking in the intermountain west; mountain biking in Moab, UT; and two nationwide recreation demand data sets aggregated by region.

Pursuant to a project meeting in Reno, NV, August 9-12, 2008, at which the hiking/mountain biking data were evaluated for completeness and consistency, the hiking data associated with evaluating the effects of fire on hiking use and value were selected for model estimation. An additional benefit of focusing on original valuation studies with a fire scenario component is the option to develop policy scenarios associated with prescribed burning for our BT exercise.

The retained data were collected using similar surveys in five states (Colorado, Montana, Wyoming, Idaho, and New Mexico). The surveys included information on past, current, and contingent recreation use at a several trails that provide an array of landscapes affected differently by past fire events. Trail attributes and socio-demographic information is also included in the data. The final pooled data restricted to hiking demand is 687 observations. This dataset is currently subjected to a final cleaning (addition of missing travel costs and fire information) and will be ready for analysis shortly.

Data Collection for Land Preservation Application

Co-PI Johnston was in charge of collecting the land preservation data. Original candidate studies for this application were drawn from 18 choice experiment analyses of WTP for farmland preservation conducted in North America between 1996 and 2007. This set includes all choice experiments from the review of Bergstrom and Ready (2005) that allow for the direct calculation of willingness to pay values per acre of farmland preserved. Additional studies, including those drawn from the gray literature, were identified through: (1) review of published research and bibliographies dealing with WTP for farmland preservation, (2) review of recent issues of resource economics journals, (3) searches of online reference and abstract databases (e.g., Environmental Valuation Resource Inventory (EVRI)) and (4) personal communication with authors known to have published research assessing farmland preservation or amenity values.[2]

Starting from this population of candidate farmland studies, which were also analyzed in the aggregate data meta-regression model (MRM) of Johnston et al. (2008), the final metadata for the individual level MRM in the present case are limited to those in the original farmland valuation metadata for which original survey data could be obtained. This involved contacting original study authors (where the PIs were not the authors of the original study) and requesting copies of original coded data. This effort resulted in 11 unique data sets, eight corresponding to CE experiments conducted in separate communities in Delaware and Connecticut in 2005 and 2007, and three County-wide experiments conducted in Delaware in 2001[3]. These datasets are ready for further analysis.

---

[2] Although the metadata include observations from all published choice experiment analysis of North American farmland preservation values, the number of such studies is limited. However, given the capacity of choice experiments to forecast welfare estimates for a wide range of preservation options—characterized by differences in multiple preservation attributes—each study in the metadata provides multiple observations of WTP.

[3] The resulting metadata are dominated by primary studies involving a small number of common authors (e.g., Johnston, Duke, Bergstrom). While this increases the possibility for certain types of selection biases in the metadata (Rosenberger and Johnston 2009), such patterns are largely unavoidable given the current composition of the relevant choice experiment literature. Nonetheless, the metadata and associated empirical results should be interpreted in light of the predominance of studies by a small number of study authors. With ongoing expansion of the choice experiment literature applied to farmland preservation, it is likely that broader samples will be available for future meta-analyses.

Model Simulations

Data generation

To gauge the econometric challenges to be expected when working with actual field data we created a simulated individual-level meta-dataset, which was then subjected to estimation via an ALMRM and an ILMRM. For simplicity and in concordance with our hiking application we casted this exercise within an Incomplete Demand System (IDS) of outdoor recreation (e.g. LaFrance and Hanemann, 1989, LaFrance, 1990, Shonkwiler, 1999, Moeltner, 2003, Hagerty and Moeltner, 2005). To start, we assume there are $s = 1 \ldots S$ sub-populations of interest, each represented by an existing source study. Each study, in turn, covers $j = 1 \ldots J$ recreation sites, and contains trip information for $i = 1 \ldots N$ individuals. For our simulated data we set S = 20, J = 6, and N = 200, which corresponds well to the typical sample size in existing recreation studies. We first generate three explanatory variables: income (*m*), travel cost (or "price") (*p*), and catch rate (*c*) as an example for a site quality indicator. These variables are drawn in a fashion that preserves the underlying hierarchical structure of "population" – "sub-population" – "individual" in their statistical properties. Summary results over studies for these simulated regressors are given in Table 1.

Second, we draw corresponding coefficients from a hierarchically structured sequence of densities. Combining these "true" parameter settings with our simulated data allows for the direct computation of seasonal trip demand and seasonal welfare. Summary results for trip and welfare statistics are given in Table 2. The first block of columns in the table summarizes results for site-specific demand, averaged over all sites and individuals comprised in a given study. The second block shows system welfare for each study, averaged over all individuals. The third set of columns depicts site-specific welfare, averaged over sites and individuals. These figures are well within the customary boundaries found in existing outdoor recreation studies. Study 11 exhibits by far the highest results for trip demand and welfare. This study could thus be interpreted as focusing on "blue ribbon" sites with respect to the activity under consideration.

Estimation

We first estimate an ALMRM, which uses only site-aggregated data. Thus, it is based on a sample of $S * J = 120$ observations. The dependent variable is the visitor-averaged compensating variation associated with access to a given site. The dependent variables are user-averaged travel cost and income, and site-specific catch rate. We estimate a semi-log model and a log-log model, using the natural log for all regressors (after aggregation). To allow for heteroskedasticity and capture intra-study correlation with a single extension (see Moeltner et al., 2007) we specify a hierarchical distribution for the catch rate coefficients. We find that the log-log model ALMRM performs well and generates predicted benefits (at the mean of all regressors) that correspond closely to sample averages (see Figure 1 and Table 3). We also estimate several ILMRMs with varying hierarchical structures, using structurally derived individual CV as dependent variable. As for the ALMRM, we consider a semi-log and a log-log form for our linear MRM.

In contrast to the ALMRM, virtually all attempted ILMRMs failed to generate benefit estimates of reasonable magnitude and range. Most suffered from serious identification problems for higher-level parameters. It became apparent that the customary linear MRM fails to capture the key features of the highly nonlinear functional form underlying the generation of the structural CV values. In other words, by switching from an aggregate to an individual-level estimation framework, we are also implicitly adopting the more complex and more highly nonlinear properties of the underlying theoretical model structure. This hampers the ability of a linear MRM to produce meaningful BT estimates, at least for our simulated application.

This crucial insight led us to a distinction between a *reduced-form* ILMRM (as considered in the original proposal) and a *structural* ILMRM, which we will examine more closely over the remaining grant period. Conceptually, the structural ILMRM would utilize the original data to directly generate welfare predictions for any desired BT context consistent with a common utility-theoretic framework. In contrast, the reduced-form ILMRM proceeds in two steps: First, raw data are used to generate welfare estimates for each individual. Second, these estimates are then regressed against a set of explanatory variables (as needed for the BT function) in a standard linear meta-regression framework. In theory, the structural ILMRM should be more efficient than the reduced-form ILMRM as it avoids the second estimation step. However, its feasibility hinges crucially on the homogeneity of outcome variables employed in the original studies. We will examine the pros and cons of either specification of the next few months.

PI Meetings

Members of the grant team met twice during this first year. The first meeting between PI Moeltner and Co-PI Rosenberger took place at the Annual Meetings of the W2133 Regional Project ("Benefits and Costs of Natural Resources Policies Affecting Public and Private Lands ") in Kona, HI, Feb. 17-20, 2008. This meeting led to a refinement of the set of candidate studies / sources for the hiking application. The second meeting between PI and both Co-PIs took place in Reno, NV, August 9-12, 2008. During that meeting the grant team took a close look at the collected data sets for both applications and identified final cleaning steps for these data. The team also discussed the simulation results and corresponding methodological issues, especially the emerging distinction between reduced-form and structural ILMRMs.

Publications / Presentations

PI Moeltner presented preliminary simulation results at the W2133 Regional Project ("Benefits and Costs of Natural Resources Policies Affecting Public and Private Lands ") in Kona, HI, Feb. 17-20, 2008. The Power point slides for this presentation are available at the PI's Project web page at http://www.ag.unr.edu/moeltner/EPA_STAR.htm. Co-PI Johnston presented preliminary results of the land preservation ALMRM at the 2008 International Workshop on Meta-Analysis in Economics and Business, Nancy, France, Oct. 17-18, 2008.

Management Issues / Changes

While project efforts have generally followed the envisioned division of labor and targeted timeline, there have been some changes in data applications, supporting personnel, and project scope. As discussed above, the originally proposed sportsfishing application was replaced with an application of meta-analysis and BT to the topic of *land preservation*. This also required a change in supporting personnel. Dr. Elena Besedin at ABT Associates, a proposed collaborator for the discarded sportsfishing application, was replaced by Dr. C.M. Starbuck (Assistant Professor, New Mexico State University) to aid Co-PI Rosenberger with the hiking application. This personnel change was approved by the Program Officer on February 26, 2008 via e-mail.

As discussed above, the scope of the Project has been broadened to some extent to consider structural vs. reduced-form ILMRMs, pursuant to the results from the simulation exercise. Our collected data will allow for the estimation of both types of models.

Budget Update

As of November 20, 2008, of the original grant amount of $198,675, $115,568 have been spent, and an additional $65,270.17 have been encumbered to cover existing purchasing orders, UNR professional salaries and benefits for the current academic year, and all indirect costs associated with these encumbrances. This leaves a balance of $17,818 which is sufficient to cover salaries and benefits for the remainder of the grant period. An itemized updated budget is available from PI Moeltner upon request.

Future Activities

For the remainder of the grant period PI Moeltner, with assistance of a graduate student, will estimate a variety ALMRMs and both types of ILMRMs for the hiking and land preservation applications. We are planning to present preliminary results at the 2009 Annual Meetings of the W2133 regional project in Austin, TX, Feb. 18-21. We are also currently preparing an abstract for a group session of the wider EPA Benefit Transfer research group at the 2009 AAEA Meetings in Milwaukee, July 26-28. We further anticipate a PI meeting at the EPA in Washington, D.C., or , alternatively, a third meeting of the grant team at a mutually convenient location to discuss estimation results and further modeling needs.

References

Bergstrom, J.C. and R.C. Ready (2005). "What Have We Learned from 20 Years of Farmland Amenity Valuation Research?" In S. Schultz, comp. Western Regional Research Publication W-1133, Benefits and Costs of Resource Policies Affecting Public and Private Land. Eighteenth Interim Report. Fargo, ND: North Dakota State University, pp. 9-36.

Hagerty, D., K. Moeltner. 2005. Specification of driving costs in models of recreation demand. *Land Economics* **81:** 127-143.

Johnston, R.J., J.B. Kukielka and J.M. Duke. 2008. Systematic Variation in Willingness to Pay for Agricultural Land Preservation: A Meta-Analysis. Presented at the International Workshop on Meta-Analysis in Economics and Business. Nancy, France. October 17-18.

LaFrance, J. T. 1990. Incomplete demand systems and semilogarithmic demand models. *Australian Journal of Agricultural Economics* **34:** 118-131.

LaFrance, J. T. and W. M. Hanemann. 1989. The dual structure of incomplete demand systems. *American Journal of Agricultural Economics* **71:** 262-274.

Moeltner, K. 2003. Addressing aggregation bias in zonal recreation models. *Journal of Environmental Economics and Management* **45:** 128-144.

Moeltner, K., K. Boyle and R. Paterson. 2007. Meta-analysis and benefit transfer for resource valuation - addressing classical challenges with bayesian modeling. *Journal of Environmental Economics and Management* **53:** 250 - 269.

Rosenberger, R.S. and R.J. Johnston. (2009). “Selection Effects in Meta-Analysis and Benefit Transfer: Avoiding Unintended Consequences.” Land Economics, in press.

Shonkwiler, J. S. 1999. Recreation demand systems for multiple site count data travel cost models. In *Valuing recreation and the environment: Revealed preference methods in theory and practice*, J. A. Herriges and C. L. Kling (eds.). Edward Elgar: Cheltenham, U.K., and Northampton MA.

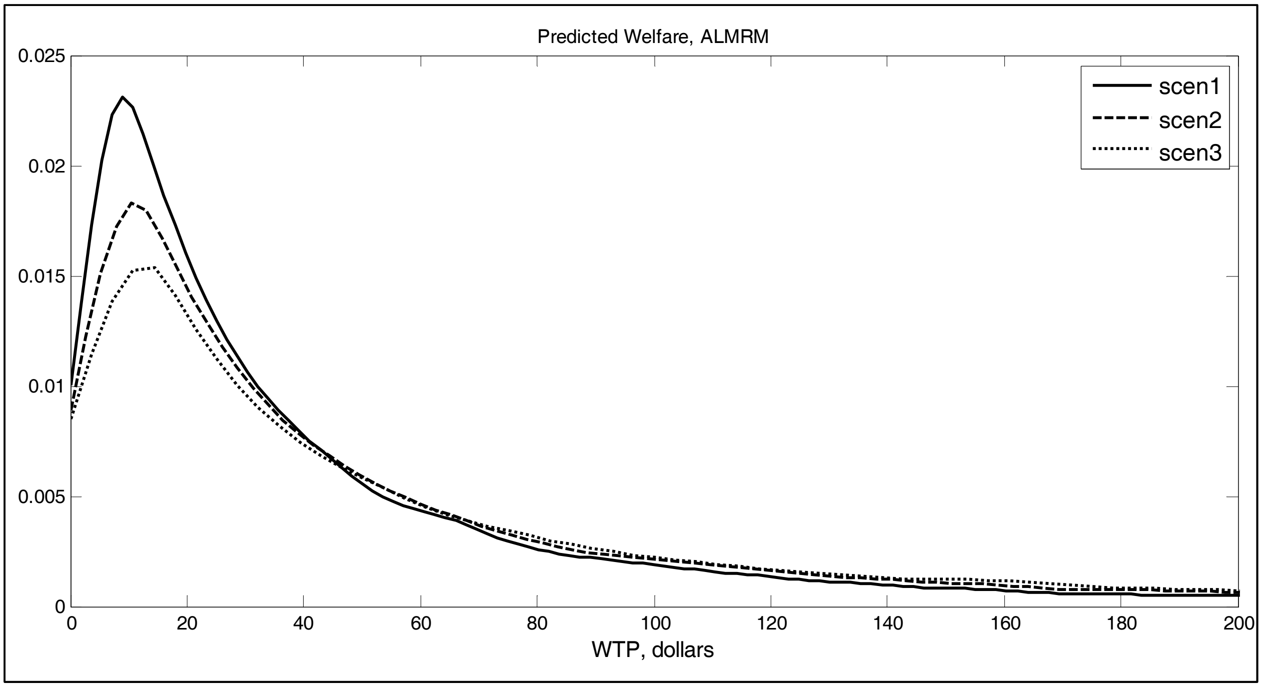
Table 1: Sample Statistics for Simulated Regressors across Studies

Study	income (000)			price			catch rate		
	mean	min	max	mean	min	max	mean	min	max
1	81.43	64.51	102.70	57.10	12.49	176.20	2.29	1.09	6.91
2	44.94	27.39	76.01	46.90	9.96	188.17	3.12	2.48	4.02
3	76.53	55.02	107.38	62.18	16.12	207.34	2.57	1.51	3.42
4	61.85	45.82	80.47	39.95	5.64	170.23	3.07	1.76	4.67
5	83.81	65.28	112.31	42.21	6.00	135.05	3.32	1.35	5.98
6	61.07	38.88	87.04	35.66	5.50	146.41	6.00	4.37	8.28
7	50.10	32.80	69.23	22.00	2.45	112.12	4.79	3.50	6.62
8	72.58	54.17	97.80	45.89	13.15	144.73	6.39	5.95	7.16
9	75.73	48.39	97.00	36.06	5.88	128.92	2.47	1.34	4.98
10	49.46	32.30	71.33	43.31	5.96	163.65	1.90	0.99	2.83
11	77.71	57.07	106.08	39.93	7.63	127.10	8.09	5.61	11.34
12	68.66	50.26	93.03	21.17	1.37	104.00	3.76	3.12	4.59
13	66.61	50.27	89.81	31.79	2.53	155.81	7.03	4.11	10.01
14	46.41	28.24	71.72	39.47	8.95	126.58	0.88	0.17	2.79
15	64.56	49.83	93.24	43.59	10.69	149.88	1.87	0.23	4.26
16	60.45	41.76	93.46	42.33	10.98	170.80	5.97	4.27	8.92
17	59.22	39.37	81.67	37.34	8.33	165.45	4.90	3.90	5.82
18	51.42	33.72	72.20	55.11	16.90	150.03	6.70	5.84	7.81
19	48.47	32.50	71.60	34.83	5.43	158.67	3.96	2.24	5.65
20	54.68	33.85	78.97	59.37	16.05	156.47	3.32	2.09	4.46

Table 2: Simulated Trip Demand and Seasonal Welfare across Studies

Study	demand per site			system welfare			welfare per site		
	mean	min	max	mean	min	max	mean	min	max
1	2.14	0.02	14.94	600.08	319.92	877.30	99.89	0.81	545.81
2	0.62	0.00	4.44	144.11	53.99	248.63	24.01	0.06	183.11
3	1.41	0.01	5.07	454.10	249.81	714.29	75.60	0.19	419.08
4	0.72	0.01	4.55	136.31	74.23	212.09	22.71	0.28	105.72
5	0.98	0.00	3.84	226.45	107.52	402.97	37.73	0.02	139.03
6	2.19	0.01	14.66	558.71	279.40	916.73	93.05	0.32	618.92
7	1.53	0.02	7.91	302.99	180.18	465.78	50.47	0.33	191.21
8	2.12	0.01	16.43	371.11	138.01	702.46	61.82	0.27	395.24
9	1.40	0.02	4.85	314.19	182.78	440.71	52.34	0.73	175.60
10	0.69	0.00	3.83	136.80	51.98	238.07	22.79	0.02	115.18
11	7.58	0.02	40.14	2,097.21	983.08	3,115.09	348.57	0.74	1,975.48
12	2.04	0.13	8.07	360.11	230.01	534.33	59.99	3.44	202.08
13	2.20	0.03	8.27	544.29	273.53	890.62	90.63	0.65	260.30
14	0.95	0.00	9.84	148.15	50.11	288.27	24.69	0.07	237.66
15	0.62	0.02	2.09	132.10	68.83	198.95	22.01	0.42	86.76
16	1.72	0.07	7.08	445.14	219.80	662.07	74.14	2.23	345.37
17	1.38	0.08	6.00	364.32	242.68	621.16	60.66	3.68	182.61
18	0.60	0.00	3.20	119.38	45.26	226.76	19.89	0.02	121.34
19	0.87	0.01	3.57	148.26	67.34	232.62	24.70	0.12	101.94
20	1.04	0.00	7.11	262.85	103.07	484.63	43.80	0.05	326.75

Figure 1: Posterior Predictive Densities for Benefits, ALMRM



Scenario 1: Travel cost, income, and catch rate set to 25th sample percentile
Scenario 2: Travel cost, income, and catch rate set to 50th sample percentile
Scenario 3: Travel cost, income, and catch rate set to 75th sample percentile

Table 3: Posterior Predictive results for the ALMRM

	mean	std	median	2.5th pct	97.5th pct
Sample	65.47	136.34	31.34	0.91	287.69
Scen. 1	51.73	70.51	26.31	2.21	269.63
Scen. 2	69.15	100.25	32.60	2.43	379.50
Scen. 3	88.22	137.63	37.99	2.37	518.79