Title: Estimating the Nonmarket Value of Green Technologies Using Partial Data Enrichment Techniques

Authors:

Brett R. Gelso
US EPA, OW
Water Policy Staff
1200 Constitution Ave.
Washington DC, 20460
202-564-5763
gelso.brett@epa.gov

Jeffrey M. Peterson
Kansas State University
Department of Agricultural Economics
331B Waters Hall
Manhattan, KS 66502
785-532-4487
jpeters@ksu.edu

Selected paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4, 2004.

Abstract: Recent studies have suggested that green technologies may be a cost effective way to manage urban runoff. Literature has also suggests that there needs to be a greater empirical basis to estimate the benefits associated with social values associated with urban trees; we therefore estimate ecosystem benefits of green technologies using emerging data enrichment valuation methods.

1.0 Introduction

Many cities throughout the United States currently face important water management issues. Among the most prominent of these issues is the control of storm water runoff from urban and nearby agricultural areas, which potentially contaminates water bodies and contributes to the risk of property damage from flooding. As an alternative to managing water flow with traditional technologies, civil engineers and urban planners are becoming increasingly interested in "green technologies." Examples of these technologies, which will both prevent flooding and mediate contamination before runoff enters water bodies, are planting trees in strategic locations and constructing wetlands. Green technologies may better be viewed as restoring natural water control methods, since urban growth replaces vegetation with unnatural impervious surfaces such as buildings and pavement.

However, as with many environmental amenities market prices may not reflect the full range of services provided from an asset. As such, the non-marketed services provided by green technologies are difficult to quantify given the virtual plethora of bias inherent in nonmarket valuation. Among the SP techniques, the Contingent Valuation Method (CVM) has evolved as perhaps the most popular of all nonmarket valuation methods but also the most controversial. Critics of CVM have pointed out several difficulties. A transparent problem is the hypothetical nature of survey responses because it is inherent in the method itself. Thus, RP approaches that rely on agents' observed behavior remain popular. One of the RP techniques is the Travel Cost Method (TCM), which elicits willingness to pay through observed decisions to visit natural sites. The obvious limitation pointed out by Hanley et al. (1997) is that explanatory variables for travel cost data may be highly collinear. Also, omitted variables result in biased parameter estimates.

Since many factors that impact travel decisions cannot be measured, travel cost models are almost always subject to omitted variable bias.

The current study attempts to adequately quantify the benefits provided by green technologies. Combined Revealed Preference (RP) and Stated Preference (SP) have emerged as a promising method to quantify environmental benefits. Such methods are thought to ameliorate respective weaknesses of each data, while taking advantage of respective strengths. And according to the Data Enrichment Paradigm, combined models offer many attractive economic and econometric properties. By adding stated data to observed data variables may be orthogonal by design, non-use values are included, statistical efficiency may be improved, and stated data is at least partially based on observed data. This paper also uniquely focuses on the role Partial Data Enrichment Paradigm when combining revealed and stated preference data in nonmarket valuation.

2.0 Background and Prior Literature

Recent literature has suggested that there needs to be a greater empirical basis to estimate the benefits from scenery, wildlife, and the social values associated with urban trees. Gelso (2002) developed a cost-minimization model that chose the optimal combination of traditional facilities, urban trees, and restored wetlands, where the non-water benefits green technologies are taken explicitly into account. Based on likely parameter values for the city of Topeka, numerical simulations of this model suggested that green technologies may be a cost effective way to manage urban runoff. Yet the results where sensitive to the amount of external benefits which were not yet adequately quantified. Although revealed preference hedonic methods may be used to estimate how much a resident will pay for scenery provided from residential trees, some research has suggested[1] that a combination of revealed and stated preference methods will more accurately account for the amenity benefits from urban trees in residential neighborhoods. As such, the objective of this paper is to estimate the nonmarket value of green technologies using combined models of revealed and stated preference data. Such 'combined' models impose the restriction that all model parameters are independent of the data source—i.e., that RP and SP data are generated from the same set of preferences. However, the hypothesis of parameter equality is not supported in many applications of conjoint analysis in nonmarket valuation[2], perhaps due to the known sources of bias inherent in these types of data. In such cases, researchers often pursue 'partial data enrichment,' when combining RP and SP data, where only certain model parameters are restricted to be equal across data sources and others are data-specific.[3] As such, we also discuss the role of the Partial Data Enrichment Paradigm when combining revealed and stated preference data.

Prior work in combining RP and SP data has center around the notion of the Data Enrichment Paradigm, where combining preference data sources complements the strengths, and reduces the weaknesses, of respective data sources. Shaikh and Larson (1998) identified the benefits of methods combining revealed and stated data. First, the SP data adds non-use values to the revealed data that are not observable in Marshallian demand equations. Second, adding stated data to the observed data improves statistical efficiency of coefficient estimates because in some cases the stated data matrix can be made orthogonal by design.[4] Finally, revealed data insures at least partially that the coefficient estimates are based on real behavior.

[1] Earnhart, D. (2001). "Combining Revealed and Stated Preference Methods to Value Environmental Amenities at Residential Locations." *Land Economics*. 77(1):12-29.
[2] Swait, J., Louviere, J. and Williams, M. (1994). "A Sequential Approach to Exploiting the Combined Strengths of SP and RP data: Application to Freight Shipper Choice." *Transportation*, 21: 135-52.
[3] Louviere, J., D. Hensher, and J. Swait. (2000). *Stated Choice Methods.* New York: Cambridge University Press.
[4] Orthogonality is where each variable is linearly independent of all other variables. A fundamental problem with revealed preference methods is that the data matrix is not orthogonal, where coefficient estimates are statistically consistent but inefficient.

Recent studies have combined RP and SP data to estimate the nonmarket value of environmental amenities. Adamowicz et al. (1994,1997) used a random utility framework to combine preferences in a discrete question format. Kling (1987) and Cameron (1992) combined discrete and continuous data using maximum likelihood. Using conjoint and hedonic price techniques, Earnheart (1997) estimated the value of environmental amenities at residential locations.

3.0 Theory of the Combined Methods

Consider the following model of consumer choice based on Louviere et al. (2000). Suppose that data was collected from the same population on observed travel decision to a resource site (RP data), as well as hypothetical decisions from a given set of alternatives (SP data). Following Cameron (1992), we assume that utility maximizing decisions of economic agents, whether from real or hypothetical data, reflect an *identical* set of preferences. Using the well-established RUM, we define our utility specifications for the RP and SP data below.

$$U_i^{RP} = \alpha_i^{RP} + \beta^{RP} X_i^{RP} + \omega Z_i + \varepsilon_i^{RP}, \forall_i \in C^{RP} \tag{0.1}$$

$$U_i^{SP} = \alpha_i^{SP} + \beta^{SP} X_i^{SP} + \delta W_i + \varepsilon_i^{SP}, \forall_i \in C^{SP} \tag{0.2}$$

where, i is an alternative in choice sets C^{RP} or C^{SP}, α's are alternative specific constants (ASCs), and X_i^{RP} and X_i^{SP} attributes common to both C^{RP} or C^{SP}, z_i and ω_i are unique to C^{RP} or C^{SP}, respectively, and the Greek letters are utility parameters. Here, the deterministic part of utility is a linear function of observable attributes for each data set, but the utility specifications do not need to be identical across data sets. This is an attractive feature because SP data may contain parameters not observable in the RP data. Notice also that the random portion of utility (ε_i^{RP} and ε_i^{SP}) are allowed to differ.

As noted, SP methods are used frequently by environmental economists to identify non-use values. For example, suppose a researcher was investigating the demand for humpback whales. Also suppose the individual collected data on the travel decision to observe the whales, as well as hypothetical choices about such trips. The given researcher may also wish to identify existence values in the SP data to provide complementary information on the demand for humpback whales. In this case, the SP data provides complementary information on utility to the RP data.

Assuming IID extreme type value 1 (EV1) error distribution between RP and SP data, with associated scale parameters (error variance) λ^{RP} and λ^{SP}, the associated models of choice for (0.3) and (0.4) are

$$P_i^{RP} = \frac{\exp\left[\lambda^{RP}(\alpha_i^{RP} + \beta^{RP} X_i^{RP} + \omega Z_i)\right]}{\Sigma_{j \in C^{RP}} \exp\left[\lambda^{RP}(\alpha_j^{RP} + \beta^{RP} X_j^{RP} + \omega Z_j)\right]}, \forall_i \in C^{RP} \tag{0.3}$$

$$P_i^{SP} = \frac{\exp\left[\lambda^{SP}(\alpha_i^{SP} + \beta^{RP} X_i^{SP} + \omega Z_i)\right]}{\Sigma_{j \in C^{SP}} \exp\left[\lambda^{SP}(\alpha_j^{SP} + \beta^{RP} X_j^{SP} + \omega Z_j)\right]}, \forall_i \in C^{SP} \tag{0.4}$$

The expressions above indicate that the probability person *i* will pick alternative *j* with *J=1,…,n* alternatives in the choice set of SP or RP data given associated utility parameters. These make up the conditional logit model (sometimes called the multinomial logit model), which allow the utility parameters to be estimated. The equations illustrate that the scale factor is a multiplicand of the utility parameters. In conventional estimation methods, the scale factor is the scale factor is confounded within the utility parameters, so that the regression coefficients are actually estimates of $\lambda^{RP}ß^{RP}$ and $\lambda^{SP}ß^{SP}$. Hence, if the error variance is not identical across equations, and if the parameter estimates across equations are identical, preferences are not identical due to the scale factor. As described in the following section, Swait and Louviere (1993) provided an appropriate method for correcting for scale differences and testing for preference equality for our purposes.

4.0 Data & Survey

Data for this analysis by interviewing 216 random patrons at sixteen parks in Topeka, Kansas. Survey questions were designed to identify amenity value of environmental services by identifying the respondent's travel costs and stated choice among Topeka parks. See Appendix I for complete survey.

Since the purpose of this analysis was to combine RP and SP data to value green technologies, the survey instrument[5] was designed to identify observed and hypothetical travel decisions to Topeka parks for respondents. Hence, the on site survey was required to obtain both observed TC data as well as hypothetical data. While many prior studies collect TC information via the mail format, Hanley et al. (2000) point out that the on-site survey format are more reliable than mail surveys since the interviewer actually witnesses the respondent's choice of resource site.

[5] The complete survey is available upon request from the corresponding author.

The following sections discuss TC and CE question formats for the survey instruments. Since the literature offers advice on optimal TC and CE formats, we follow closely the experimental designs of prior studies.

The Travel Cost Model

Although the TCM studies are ubiquitous in the literature, few studies entirely address the problems of substitute sites, multiple purpose visits, and opportunity cost of time. Our survey attempts to control these issues through our detailed questionnaire format. As well as observed park choice, the other parks the respondent chooses to visit. In this vein, choice set is used to describe the parks that are available to respondent in his/her respective area. Also, we determine the agent's opportunity cost of leisure time through standard calculations in Fugitt and Wilcox (1999).

Another challenging aspect of the TCM emerges when the respondent have numerous resource sites available. Such large choice sets may result in an intractable model. Parsons and Kealy (1992) show that relatively small choice sets for each person can be constructed by including the chosen alternative(s) and as few as three randomly drawn rejected alternatives. This method results in a model that yields consistent parameter estimates. Our study adopts this approach. In particular, each respondents choice set contains the parks actually visited plus three additional randomly selected parks.

Conjoint Experiment

For our purposes, we use the hypothetical CE to mirror the observed travel decision of the respondent. While prior studies have combined conjoint data with the travel cost decision, our study adopts a unique CE approach in order to more closely complement the SP and RP travel decision of respondents. We include the respondent's intensity of preferences for Topeka parks, as well as choice of parks. This follows recent literature that has incorporated the consumer's intensity of preferences in CE. Kuperis et al. (1999) used CE to the demand for milk products by designing stated choice experiments to identify choice of milk, as well as the quantity demanded for each choice of milk. Figure 1 (Appendix) illustrates that the respondent is faced with several resource allocation decisions for park sites. For example, a given consumer was told that there were only three parks in his/her area and asked to choose how often he or she would visit each park with certain amenities.

The CE offers several benefits to our analysis of combining RP and SP data. First, this approach mirrors the observed travel decision for respondents. That is, in the observed data the agent chooses which site to visit as well as the number of times to visit. Hence, the CE data exactly reflects the observed travel decision.

As discussed in an earlier section of this paper, hypothetical data is beneficial to researchers since the data matrix may be constructed orthogonal by design. In the literature this is called the Orthogonal Main Effects experimental design, where every respondent may be given a unique set of alternatives for the given choice model. In our case we design seventy-two surveys, where each respondent chooses from unique combination of parks amenities.

In the next section we first describe the estimation of separate TC and CE models. Preferences are specified in a random utility framework, where the deterministic component of utility depends on the urban amenity, and a set of demographic variables.

Next, data is combined through the vertical concatenation of stated and revealed preferences in the multinomial logit (MNL) model of choice. Subsequent hypothesis tests are performed regarding the preference equality across utility parameter in SP and RP preference choice models. We also estimate the relative scale factor, and estimate how much of the RP error variance is that of the SP data. That is, an interesting part of the combined models to investigate how much SP and RP data actually reflect the same utility functions. According to Cameron (1992), SP and RP data theoretically reflect an identical set of preferences. However, as with many empirical methods, achieving the nexus between theoretical economics and applied methods is indeed challenging.

Since the purpose of this analysis was to combine RP and SP data into a single model to predict park choice, the survey instrument was designed to identify observed and hypothetical travel decisions to Topeka parks for respondents. Hence, the on site survey was required to obtain both observed TC data as well as hypothetical data. While many prior studies collect TC information via the mail format, Hanley et al. (2000) point out that the on-site survey format is more reliable than mail surveys since the interviewer actually witnesses the respondent's choice of the resource site.

5.0 Econometric Analysis

In this section we analyze respondent's choice of resource site to identify the non-market value of the park amenities. Using the random utility framework, "MLE techniques are used to identify the deterministic components of utility in the MNL model of choice (Earnhart, p. 19)."

Structure

Before discussion of separate and combined models of choice, several considerations are necessary to identify. The variables included in both models are listed and defined in Table 1. First, in order to identify the demand for park amenities, 1,0 dummies are used for high or low tree density, as well existence of a water feature, garden, athletic field, and playground. As mentioned earlier, the stated data adopted the Orthogonal Main Effects experimental design such

Table 1: Summary Statistics and Variable Definition

Variable	Description	Stated Data	Revealed Data	Combined Data
I. Park Amenities:				
Athletic Field	if present (=1), otherwise (=0)	0.484 (0.499)	0.966 (0.182)	0.580 (0.493)
Water Feature	if present (=1), otherwise (=0)	0.459 (0.499)	0.648 (0.478)	0.497 (0.500)
Tree Density	high (=1), low (=0)	0.496 (0.500)	0.393 (0.489)	0.476 (0.500)
Garden	if present (=1), otherwise (=0)	0.538 (0.498)	0.637 (0.482)	0.558 (0.497)
Playground	if present (=1), otherwise (=0)	0.493 (0.500)	0.989 (0.106)	0.592 (0.491)
II. Demographic Characteristics Interacted with Price:				
Education	years of education	15.131 (2.66)	15.381 (2.620)	15.181 (2.658)
Sex	gender (1=Male)	0.406 (0.491)	0.358 (0.481)	0.396 (0.489)
Adults	number of adults	1.920 (0.627)	1.897 (0.706)	1.915 (0.644)
Children	number of children	1.320 (1.215)	1.225 (1.064)	1.301 (1.186)
Urban	residential location (1=Urban)	0.783 (0.412)	0.805 (0.396)	0.787 (0.409)
Income	dollars of income per annum	51285.71 (33785.11)	52681.30 (33052.56)	51564.41 (33632.23)
Age	*years of age*	*30.360* (10.059)	*30.616* (9.310)	*30.411* (9.911)
III. Price of Resource Site and Observed Park Attributes				
Price	price of travel plus opportunity cost of leisure time	8.98 (6.37)	17.26 (16.506)	10.64 (9.89)
IV. Characteristics Unique to Observed Parks				
Center	presence of community center	---------	0.293 (0.456)	0.058 (0.235)
Gage	very large park	---------	0.344 (0.475)	0.0686 (0.252)
Small Park	small park	---------	0.003 (0.061)	0.0007 (0.027)
Size	size of park in acres	---------	202.236 (268.126)	40.385 (144.408)
No. of Observations		1050	262	1312

that each respondent faces a park with a unique set of amenities. The design produces a data matrix that is orthogonal, where each variable is linearly independent of all other variables. For the observed travel cost data, binary dummies are also associated the respective park features. Second, price is interacted with demographic feature of respondents. This interaction is required since price is the only feature that varies among individuals (Earnhart, p. 19).

Estimation

This section illustrates MLE procedures used to identify the demand for urban park amenities with separate revealed and stated models of choice, as well as the combined model. All models were estimated by maximum likelihood using LIMDEP Econometric Software Version 7.0.

The following sections proceed as follows. First, separate SP and RP models are illustrated and interpreted. The purpose of the section will be to investigate the credibility of the data enrichment hypothesis. A graph of parameter estimates will provide preliminary assistance for determining which parameters are based on the same utility functions. Next, the joint estimation results are presented as well as the test statistic for data enrichment.

Separate SP and RP Model Results

Results for the multinomial regression for the separate and combined models are presented in Table 2. We first compare the coefficient estimates of the separate RP and SP choice models. Several coefficients are similar in sign and magnitude, while many coefficients are quite dissimilar in sign and magnitude. The parameters WATER FEATURE, GARDEN, PALYGROUND, URBAN, INCOME, and PRICE have the same sign in both models. However, the coefficients for TREE DENSITY, EDUCATION, SEX, ADULTS, CHILD, and AGE yielded opposite signs. Also, even though PLAYGROUND is the same sign in SP and RP results it is dissimilar. That is, most parks in the observed preference data had a playground and it is hypothesized that this lack of variation resulted in a sign reversal. Similar to many studies with RP data, the magnitude appears to be overstated on the coefficient for PLAYGROUND due to lack of variation in the data.In terms of data enrichment, this data suggests that some, but not all utility parameters are comparable between revealed and stated preference models.

Interesting results emerge from the separate SP and RP models of choice. First, many variables in the models are statistically significantly. However, the SP choice model displayed less robust results with respect to statistical significance of demographic variables. Although Louviere et al. (2000) indicated that sign reversals generally occur in RP data, our results suggest that the RP data was more robust compared to the SP data. Our hypothesis is that the lack of statistical significance is due in large part to the data collection process (i.e., the on-site survey).

We now seek to test the hypothesis of preference equality between RP and SP parameters, using the visual test first used by Swait and Louviere (1993). As noted by the authors, the test only provides initially evidence for data pooling since the plot does not include sampling errors. As described in an earlier section, in order to test for preference homogeneity we must first relax the IID/EV1 assumption of constant error variance. Relaxing the assumption means that the estimated parameters are confounded with the scale factor. Even if estimated parameters are identical between data sets, results may not be identical since our scale parameter is a multiplicand of the coefficient.

The hypothesis of data enrichment may now be explored with the estimated coefficient from the MNL model of choice. According to Louviere et al. (2000), a preliminary test for data enrichment is to plot the RP against the SP parameter estimates (Figure 2).

Combined SP and RP Model Results

Results for the combined SP and RP model are presented in Table 2. All coefficient estimates are significant at the 0.01 level. The psuedo r-square indicates a very high goodness of fit for the model. Perhaps the most interesting interpretation of model results is the estimated value of the scale factor of 0.30.

The value of 0.40 indicates that 16% of the error variance in the RP preference is that of the SP preference data. Hence, if preference equality exists then the data must be rescaled such that parameter estimates are comparable.

The question that now remains is whether we can accept the hypothesis of data enrichment. Using the procedure from Swait and Louviere (1993), we use the test statistic $2[(L^{RP}+L^{SP}]-L^{Joint}]\sim\chi_{\beta-1.}$ The associated chi-square statistic with 14 degrees of freedom in our combined model is heavily in disfavor of the hypothesis of parameter homogeneity between data sets.

Our results yield tentative evidence for data enrichment. The plots for TREE DENSITY, EDUCATION, SEX, ADULTS, CHILD, and AGE are in the unexpected area of the parameter plot. Louviere et al. (2000) suggested that such results may warrant pursuing "Partial Data Enrichment", such that only particular utility parameters are identical between RP and SP preference data. The next section explores partial data enrichment.

Table 2: Multinomial Logit Regression Results

Variable	Stated Data	Revealed Data	Combined Data	Partially Combined Data
Athletic Field	0.422***	---------	0.722***	0.212_{SP}***
	(0.018)		*(0.059)*	*(0.008)*
Water Feature	0.374***	0.0924*	0.603***	0.184_{SP}***
	(0.018)	*(0.055)*	*(0.033)*	*(0.008)*
Tree Density	0.176***	-0.345***	0.328***	0.079_{SP}***
	(0.195)	(0.066)	(0.036)	(0.009)
Garden	0.224***	0.055	0.225***	0.111_{SP}***
	(0.018)	*(0.080)*	*(0.037)*	*(0.009)*
Playground	0.550***	3.750***	1.412***	0.271_{SP}***
	(0.017)	*(0.305)*	*(0.040)*	*(0.008)*
Education	-0.022***	0.009***	0.007***	0.010_{RP}***
	(0.003)	*(0.001)*	*(0.001)*	*(0.0005)*
Sex	0.028**	-0.074***	-0.065***	-0.065_{RP}***
	(0.015)	(0.006)	(0.006)	(0.005)
Adults	0.105***	-0.0004	0.007 **	0.043_{SP}***
	(0.011)	(0.004)	(0.004)	(0.004)
Children	-0.0003	0.021***	0.017***	0.015_{RP}***
	(0.005)	(0.002)	(0.002)	(0.002)
Urban	-0.004	-0.065***	-0.059***	-0.037_{C}***
	(0.017)	(0.008)	(0.007)	(0.005)
Income	-0.01E-05	-0.581E-06***	-0.556E-06***	-0.511E-06_{C}***
	(0.02 E-05)	(0.830E-07)	(0.824E-07)	(0.642E-07)
Age	-0.08E-02	0.002***	0.001***	0.0007_{C}***
	(0.07E-02)	(0.0003)	(0.0003)	(0.0002)
Price	-0.009	-0.096***	-0.076***	-0.112_{C}***
	(0.050)	(0.022)	(0.02)	(0.013)
Size	---------	0.0003***	0.0002***	0.0009_{RP}***
		(0.614E-04)	(0.572E-04)	(0.523E-04)
Center	---------	-0.351***	0.258***	-0.038_{RP}
		(0.051)	(0.041)	(0.301)
Gage	---------	0.840***	0.093**	0.896_{RP}***
		(0.075)	(0.047)	(0.029)
SmallPark	---------	-1.913***	-1.582***	-1.723_{RP}***
		(0.074)	(0.071)	(0.069)
No. of Observations	350	167	517	517
Log-Likelihood at Zero	-23345.51	-21150.4838	-67841.5060	-67841.5060
Log-Likelihood at Convergence	-22117.14	-13615.72	-36427.52	-36286.66
McFadden's ρ^2	0.05262	0.35625	0.46305	0.46513
Estimated Scale Factor λ			0.40	2.00
Test Statistic for Data Pooling:			1389.32	43.15
$2[(L^{RP}+L^{SP})-L^{Joint}] \sim \chi^2_{\beta-1}$			Accept=No	Accept=No

Partial Data Enrichment

Following the advice of prior literature, this section combines data through the so-called partial data enrichment process. Louviere et al. (2000) suggested that that removing certain parameters from the combined model may result in acceptance of the hypothesis of data enrichment. If coefficients in the visual test are in quadrants I and IV, or are far away from the cloud of points, the authors recommended removing such parameters from the joint model.

Given the partial data enrichment process, we now consider the appropriate coefficients that had the same sign and similar magnitude to be combined. The purpose of the models is to examine the consequences of excluding explanatory variables in the pooled models. Prior work has suggested the hypothesis of data enrichment may be retained with partially enriched models of choice.

Louviere et al. (2000) offers certain advice for researchers who wish to partially enrich data. First, the authors suggested that sign reversals only occur very rarely in the SP data. However, in our case we observe that the RP data gives much more reasonable coefficient estimates compared to the SP data. Also, the goodness of fit in the RP model (see Table 2) has much higher goodness of fit compared to the SP model. Second, the author indicated that there is differential relative importance of coefficient estimates in RP and SP models of choice. For example, suppose that the coefficient estimate for CHILD in the separate SP and RP models were opposite in sign (See Table 2). In this case the sign for CHILD in the SP data was negative, but the sign for CHILD was positive in the RP data. Clearly it is reasonable to assume that the more children an individual has the more the person will visit a given park. Hence, CHILD is included in the forthcoming partially combined models but only through the RP data, and the variable is omitted in the SP data.

Louviere et al. (2000) suggested that the process of choosing which variables to include in the partially enriched model depends on the experience of the researcher and economic theory. The author also indicated that few studies have addressed which variables to include in partially enriched models, and that this is a subject that needs to be addressed in future research.

In our partially enriched model, we combine explanatory variables with the same sign and magnitude, but also variables that exist in regions I and IV of the parameter plot. However, the variables in regions I and IV were only included in the respective SP or RP data with reasonable signs. Hence, the data matrix for the partially combined model is given as:

$$Q(\lambda^{SP}) = \left(\left| \begin{array}{c} X^{RP} \\ \lambda^{SP} X^{SP} \end{array} \right| \begin{array}{c} Z^{RP} \\ 0 \end{array} \left| \begin{array}{c} 0 \\ \lambda^{SP} W^{SP} \end{array} \right. \right) \tag{0.5}$$

where Z^{RP} and W^{SP} are variables included in the respective RP or SP part of the data matrix and all other terms are previously defined.

In general, the RP travel cost data behaved much better than the SP data. In most cases, coefficient estimates were the expected sign and were consistent with economic theory. Hence, the variables EDU, SEX, and CHILD were included in the RP portion of the combined model. Also, the travel cost data is thought to have popular playgrounds. Hence, it is thought that the coefficient is overstated in the RP data. In our partially combined model the utility coefficients for URBAN, INCOME, AGE, and PRICE were combined since separate models suggested these parameters were the same in sign as well as magnitude. Results for the partially combined model are presented in the fourth column of Table 2.

Interpretation of Partially Enriched Models

Interestingly, relatively few studies have addressed the issue of partial preference homogeneity in combined RP and SP models. Swait et al. (1994) used a transportation model to partially combine data sources. The authors found that the hypothesis of data enrichment was rejected due to several particular parameters that exist in regions I or IV of the visual test. Once the suspect parameters were discarded the relative scale parameter improved from 0.708 to 0.941. Recalling the interpretation of the scale parameter from (0.20), the result is the amount of error variance that is common between the data sets. Hence the amount of error variance that was identical in the authors study improved from 50% to a robust 86%.

Such results are hopeful for researchers enthusiastic to combine RP and SP data. However, the critical difference in our study is that the preference data is that of non-market goods that are not normally traded in the marketplace. Clearly analysts face numerous biases with non-market valuation, and such biases make the data enrichment hypothesis less likely to accept.

The partially enriched model, with certain variables isolated in the respective RP and SP, are presented in Table 2. Recall that variables in the model were partitioned in respective RP and SP models depending in if they had displayed the expected sign in the initial fully enriched model. Even in the case of partial enrichment the hypothesis of data enrichment is rejected even at the 1% level of significance, suggesting the utility functions in the empirical model are quite dissimilar.

The chi-square statistic $-2[(L^{RP}+L^{SP}]-L^{Joint}]\sim\chi_{\beta-1}$ is 21.0261 (14 d.f.) and the critical value α=0.05 level is 43.15, which again indicates that we should reject the hypothesis of data enrichment. However, the reader may notice the test statistic is much improved compared to the fully combined model.

The reader also wish to test his/her understanding of the benefits of combined models. Notice that the fully and partially combined models in Table 2 display quite robust statistical results. The psuedo r-square is not only very high, but improved a great deal compared to the separate models. Also, the fully and partially combined model yields statistical significance in almost all parameters at the 99% level as well as low standard errors.

6.0 Welfare Analysis

Estimating the Value of Green Technologies

Table 3 presents the welfare effects for the separate and combined models. The Compensating Variation (CV) of income is estimated for the Green Technologies at Topeka parks. As noted, the RUM is used to identify resource site choice in SP, RP, combined, and partially combined models. Consider the CV for the MNL model of choice:

$$CV = \frac{-1}{\mu}\left[\ln\sum_{i=1}^{n} e^{V_i^0} - \ln\sum_{i=1}^{n} e^{V_i^1}\right] \tag{0.6}$$

where μ is the marginal utility of income, the price coefficient from the estimated regression.

In general, reasonable estimates were obtained from the RUM. However, the SP data produced an unexpected large CV for both WATER FEATURE and TREE DENSITY. Such overstatement of willingness in SP surveys is consistent with prior research, since respondents are not bound by a budget constraint and may overstate or understate true preferences. The sign reversal for TREE DENSITY (i.e., CV<0) in the RP model indicates that individuals obtained disutility of ta

TABLE 3: Per Trip Welfare Calculations for Multinomial Logit Regression Results of Stated and Revealed Preference Data with Randomly Drawn Sets: Estimating the Value of Green Technologies

Natural Feature	Stated Data	Revealed Data	Combined Data	Partially Combined Data
Water Feature	$21.57	$0.59	$4.55	$0.94
Tree Density	$9.46	$-2.53	$2.56	$0. 42

The above table presents per trip CV for the SP, RP, Combined, and Partially Combined RUM models of choice. In general, reasonable estimates were obtained from the RUM. However, the SP data produced an unexpected large CV for both WATER FEATURE and TREE DENSITY. Such overstatement of willingness in SP surveys is consistent with prior research, since respondents are not bound by a budget constraint and may overstate or understate true preferences. The sign reversal for TREE DENSITY (i.e., CV<0) in the RP model indicates that individuals obtained disutility of income from the presence of high TREE DENSITY as a result of the negative coefficient. However, this type of problem commonly occurs in RP data due to lack of variation in data (Louviere et al., 2000).

income from the presence of high TREE DENSITY as a result of the negative coefficient. However, this type of problem commonly occurs in RP data due to lack of variation in da (Louviere et al., 2000). Such results illustrate the usefulness of the combined models. Even though hypothesis of data enrichment was rejected in the combined and partially combined models, results suggest that welfare calculations are more reasonable compared to separate SP and RP models.

7.0 Conclusions

In this study, our primary benefit from the pooled models is thought to be the improvement of statistical efficiency and robustness of regression results. Furthermore, economic theory tells us that the RP data should have the same underlying utility sets as the SP data (Cameron). However, examining the results of our data may suggest that the connection between economic theory and empirical studies is challenging.

The models presented in this paper indicated rejection of the hypothesis of data enrichment. However, the rejection of data enrichment is not surprising when compared to recent studies. Similar studies have also combined RP and SP preference data in a very similar metric conjoint approach, finding sufficient evidence to reject the data enrichment hypothesis.

Although data enrichment was rejected in this paper, the results offer interesting implications for non-market valuation. First, an interesting result of this analysis is the exploration of the Partial Data Enrichment Paradigm. Few studies have addressed partially combining preference data for environmental goods. Although we reject partial data enrichment, future researchers may wish to also test the hypothesis.

Due to the virtual plethora of forms of bias, it is clearly challenging to combine SP and RP data for environmental goods. The rejection of data enrichment may serve to highlight the obvious problems with non-market valuation. From economic theory it is understood that RP and SP have an identical underlying set of utility preferences. However, in practice, the nexus between economic theory and applied analysis is indeed less transparent.

Extensions of this study are numerous and interesting. Recent literature has combined revealed and stated preference data with RUMs to estimate the value of environmental amenities, noting a number of economic and statistical benefits. Shaikh and Larson (1998) discussed that a clear problem is the RUM model assumes the multinomial logit functional form. More importantly, the RUM method is not a classical demand system[6] with theoretically appealing demand restrictions whereby elasticities may be obtained. The authors combined Contingent Valuation and Travel cost data using an AIDS model, noting benefits of using flexible functional forms when combining revealed and stated reference data.

Interestingly, Shaikh and Larson (1998) also discussed that future research in combined methods could compare RUMs with a classical demand system. Future research may adopts this approach. Using SP and RP data from our study an interesting extension study would be to

[6] Demand systems may also correct for contemporaneous correlation and improve statistical efficiency. See Griffiths et al. (1993).

combine the data in both the RUM and AIDS models. Such research is currently underway by the authors.

8.0 Works Cited

Amadowicz et al. (1994). "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management.* 26: 271-292.

Amadowicz et al. (1997). "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Methods of Environmental Valuation." *Journal of Environmental Economics and Management.* 32: 64-84.

Cameron, T. (1992). "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics*. 68(3), 302-17.

Earnhart, D. (2001). "Combining Revealed and Stated Preference Methods to Value Environmental Amenities at Residential Locations." *Land Economics.* 77(1):12-29.

Fugitt, D. and S. Wilcox. (1999*). Cost-Benefit Analysis for Public Sector Decision Makers*. London, Quorium Books.

Gelso, B. (2002). *Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Green Technologies.* Doctoral Thesis. Kansas State University. Dept. of Agr. Economics.

Green, W. (1998). *Limdep User's Manual: Version 7.0*. Plainview, New York: Econometric Software, Inc.

Hanley, N. et al. (1997), *Environmental Economics*. New York: Oxford University Press, Inc. , pp. 404-409.

Kling, C. (1997). "The Gains from Combining Travel Cost and Contingent Valuation Data to Value Nonmarket Goods." Land Economics. 73(3):428-39.

Kuperis et al. (1999). "Consumer's Responses to the Potential Use of Bovine Somatotrophin in Canadian Dairy Production." *Canadian Journal of Agricultural Economics.* 47: 151-163.'

Louviere et al. (2000). *Stated Choice Methods.* New York: Cambridge University Press.

Parsons, G. and M. Kealy. (1992), "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics.*68 (1): 93-106.

Shaikh, S. and D. Larson (1998), "Methods for Combining Travel Cost and Contingent Valuation Data." Unpublished paper prepared for the World Congress of Environmental and Resource Economists, Venice, Italy, June 25-27, 1998.

Swait, J., Louviere, J. and Williams, M. (1994). "A Sequential Approach to Exploiting the Combined Strengths of SP and RP data: Application to Freight Shipper Choice." *Transportation*, 21: 135-52.

Swait, J. and J. Louviere. (1993). "The Role of Scale Parameter Estimation in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research*, 30: 304-314.

9.0 Appendix

Figure 1: Example of Intensity of Preferences and Orthogonal Main Effects of Conjoint Experiment

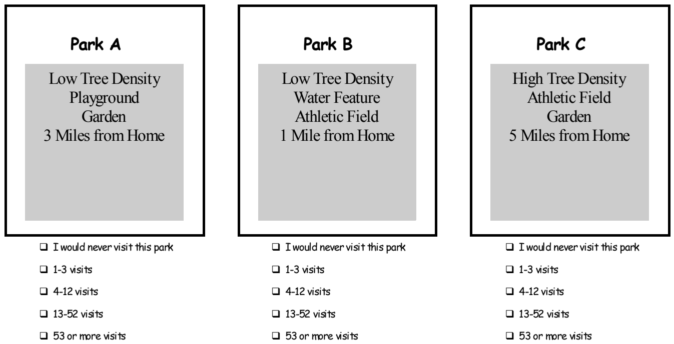

Figure 1: The diagram illustrates the CE experimental design for this analysis. Each respondent faces a unique choice set of three parks with 2^5 combinations of park amenities. The so-called Orthogonal Main Effects design results in a data matrix that greatly reduces multicollinearity. As mentioned earlier in this analysis, the benefit of stated preference data is that it may be designed to have several desirable statistical qualities such as low collinearity and standard errors. Conversely, observed data for environmental explanatory variables are often highly related and therefore collinear. As such, if preference equality exists, the stated data may be combined with the observed data to result in parameter estimates that are statistically efficient. Indeed, statistical efficiency is a desirable quality to researchers, as it provides more stable and efficient parameter estimates.

Yet another desirable quality of our conjoint data is the inclusion of intensity of preferences for parks to mirror the observed travel resource allocation. While prior studies have combined travel cost and conjoint data, our study uniquely combines that that includes intensity of preferences in both observed and hypothetical choice of Topeka parks.

Figure 2: Visual Test for the Preference Equality of Estimated Coefficients for SP and RP Data

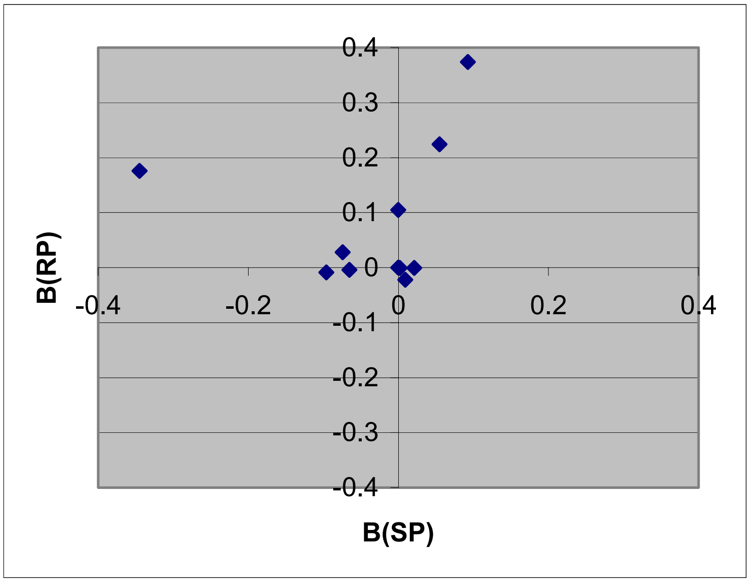

Figure 2: The figure above illustrates the visual test for preference equality across RP and SP preference data. In order to test for preference equality, we are testing $\lambda^{RP}\beta^{RP} = \lambda^{SP}\beta^{SP}$. Rearranging this expression, we may obtain the expression $\beta^{RP} = \lambda^{SP}/\lambda^{RP}(\beta^{SP})$, where the slope of this function is $\lambda^{SP}/\lambda^{RP}$. Using the estimated coefficients for RP and SP data from Table 1, the RP coefficients are plotted against the SP coefficients. According to Louviere et al. (2000), estimates in quadrants II and IV are preliminary evidence of preference equality across respective utility sets. However, since relaxing the IID/EV1 assumption results in parameter estimates that are a multiplicand of the scale factor such that $\lambda^{RP}\beta^{RP} = \lambda^{SP}\beta^{SP}$, the plot only contains information regarding the equality of utility parameters and entirely ignores error variance.

Interestingly, while the parameters PRICE , INCOME, URBAN, WATER FEATURE, GARDEN, and TREE DENSITY are in quadrants II and IV, the coefficients for TREE DENSITY, EDUCATION, SEX, ADULTS, CHILD, and AGE are not in expected regions. Also, PLAYGROUND exists far outside the cloud of points in quadrants II and IV. Louviere et al. (2000) suggests that this data indicates that utility functions are not equal across data sets. However, the outliers may be discarded in order to pursue “Partial Data Enrichment” such that only particular parameters reflect the same utility preferences in RP and SP preference data.