Does economic endogeneity of site facilities in recreation demand models lead to statistical endogeneity?[1]

Min Chen[2]

Michigan State University

Frank Lupi

Michigan State University

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2009

AAEA & ACCI Joint Annual Meeting, Milwaukee, WI, July 26-28, 2009

[1]Previously titled: "Implications of Endogenous Site Characteristics in Random Utility Models"
Control ID: 613614

[2]Corresponding author: Dept. of Agricultural, Food and Resource Economics, 208 Cook Hall, Michigan State University, East Lansing, MI 48824, USA. Ph: 517-352-7193. E: chenmin5@msu.edu

Does economic endogeneity of site facilities in recreation demand models lead to statistical endogeneity?

Abstract

Different kinds of endogeneity problems in Random Utility Models of recreation demand have been studied in previous literature. Some site characteristics, like facilities, could be endogenous in an economic sense due to the interplay of supply and demand. That is, it may be that more popular recreation sites tend to have better site characteristics since managers with limited budgets would be more willing to invest in them. If recreation site improvements are more likely to occur at the more popular sites, then might this economic endogeneity cause problems for econometric models linking site demand to facilities. In this paper, we use Monte Carlo simulations to test whether this economic endogeneity will lead to statistical endogeneity.

Keywords: Random Utility Models; Facilities; Endogeneity; Monte Carlo simulations

Acknowledgements

The authors thank the Michigan Experiment Station, the Michigan State University Environmental Science and Policy Program, and the NOAA Coastal Stressors program for support of this research. The authors are solely responsible for the contents.

1 Introduction

Random Utility Models (RUMs) are widely applied in the travel cost technique for valuing recreational activities, relating visitation to travel costs and site characteristics. Discrete response models, like multinomial logit or conditional logit, are used to estimate people's choice behaviors. From the econometric standpoint, obtaining consistent estimates requires the exogeneity of the independent variables like travel costs and site characteristics.

Specification problems potentially causing bias in travel cost methods were paid attention to as early as 1970s, especially the omission of travel time variable and congestion effects. Cesario and Knetsch (1970), Brown and Nawas (1973) and Gum and Martin (1975) discussed how to incorporate travel time and reduce its multicollinearity with travel cost at the same time; McConnell and Duff (1976) and Wetzel (1977) stated that congestion effects, if there were any, should be incorporated into the travel cost model to avoid estimation bias. Allen Stevens and Barrett (1981) found that the impact of excluding travel time and congestion varied from situation to situation. Caulkins, Bishop and Bouwes (1985) showed that the omission of cross-price variables did not necessarily cause bias, and the sign of the omission bias was determined by the true economic relationship.

Recent studies have been focusing on the possible endogeneity in RUMs. Following Ben Akiva and Leerman (1985), Haab and Hicks (1997) raised the issue that the set of alternatives, rather than defined by researchers, could be endogenously determined by individuals. They added weighted probabilities to the log likelihood function to reflect the probability that certain sites are selected into the set of alternatives, and the estimation results turned out to be very different. Murdock (2006) studied unobserved site characteristics, which were absorbed into the error term, which could be correlated with the travel cost variable. Monte Carlo simulations were used to test whether the proposed approach for addressing this endogeneity problem performed better than the traditional methods. Timmins and Murdock (2007) stated that the omission of the variable of congestion in the estimation would lead to significant endogeneity problems, since it depended on real visits. They supposed individuals made rational decisions given others' choices and considered Nash equilibrium in repeated games. A quantile regression with instrumental variables was applied to get new estimates. Von Haefen and Phaneuf (2008) developed a combined revealed and stated preference approach to overcome the endogeneity of unobserved determinants.

Those endogeneity problems addressed in this literature have mainly focused on the site selection, congestion and omitted variables, and are corrected to ensure the consistency of estimates. Now, let's consider site characteristics, for example, facilities. Many studies have found that facilities variables are often significant in explaining people's recreational behaviors. Parson (2003) reported the presence of amusement parks and restroom facilities as explanatory variables in the latent utility equation, and their estimated coefficients were statistically significant at 95% level of confidence. Lew and Larson (2005) included lifeguard presence and parking availability dummies as two explanatory variables for beach use, which were also statistically significant. Von Haefen, Massey and Adamowicz (2005) used bathroom availability and public parking in their recreational demand estimation. Yeh, Haab and Sohngen (2006) took into account the effects of lifeguard and number of picnic tables when valuing recreation trips to beaches. Cutter, Pendleton and DeShazo (2007) considered the effects of toilets, trails, tables and benches in their model of recreational demand.

At the same time, the supply and types of facilities are also determined by people's visitation as the literature in parks management makes clear. Lee and Driver (1999) compared three recreation resource management frameworks: activity-based management (ABM), experience-based management (EBM) and benefits-based management (BBM). BBM is an extension of the first two, aiming at providing public recreation opportunities which people benefit from. Shin, Jaakson and Kim (2001) pointed out that "Benefits-based management seeks to provide recreation benefits for recreation participants by managing the physical environments in which recreation occurs", and they included facilities and their maintenance as one attribute of the setting of recreational sites. Faghri, Lang, Hamad and Henck (2002) mentioned a set of criteria for where to optimally locate park-and-ride facilities, one of which suggested that a site with lots of traffic passing through should be a suitable location. Cook (2008) used a benefit transfer method to estimate the value of a new long-distance walking trail in a tropical rainforest. If no people went for recreational activities in the forest, managers would not be likely to build a walking track since its value was low.

If we view the managers as the supply side and the recreationists as the demand side, managers change facilities in response to recreational demand, and recreational demand varies in response to facilities. The interplay of supply and demand makes facilities endogenous in the economic sense. Usually, this will cause inconsistency of estimates in econometric models, but not necessarily. Therefore, the objective of this paper is to examine the extent to which economic endogeneity of facilities leads to statistical endogeneity. If it does, we should use instrumental variables to address this problem; if it does not, we don't have to worry about it.

To address the issue, Monte Carlo simulations are applied. In the simulations, we set values for the "true" parameters, simulate choice sets, run regressions, and obtain estimates. If estimates converge to the true parameters, they are consistent and the economic endogeneity of facilities does not matter. If they do not converge, then facilities are statistically endogenous. The advantage of Monte Carlo simulations is that we know what the "truth" is; otherwise, with empirical data, we can test the statistical endogeneity, but we cannot judge the consistency of a certain estimator for sure without knowing the true values.

In the following sections, we present the basic choice model for our recreation demand simulations. In the simulations, we first assume all explanatory variables including facilities are exogenous to test that the approach works for the base case. Next, we let facilities be determined by recreational demand and supply, and investigate whether we still get consistent estimates under this circumstance. Then, we conduct sensitivity analysis, changing the underlying factors of simulations. Finally, the results from simulations are discussed.

2 Methods

2.1 Conditional Logit Models

In RUMs, the latent utility that person i gains from visiting site j is:

$$U_{ij} = X_{ij}\beta + \varepsilon_{ij}$$

Where X_{ij} includes travel cost, which varies across people and sites, and site characteristics, which only varies across sites; ε_{ij} is a random term counting for unobserved preferences. If there are J sites and individual i chooses to go to site k, it must be that:

$$U_{ik} = \max\{U_{i1}, U_{i2}, \dots, U_{iJ}\}$$

The revealed choice variable for this person would be a set of binary responses:

$$(y_{i1}, y_{i2}, \dots, y_{ik}, \dots, y_{iJ}) = (0,0, \dots, 1, \dots, 0).$$

Following McFadden (1974), when ε_{ij} follows a Type I extreme value distribution, the maximization of the random utilities yields site choice probabilities given by a conditional logit model where the probability that individual i chooses site k is:

$$Pr_i(k) = \frac{e^{X_{ik}\beta}}{\sum_{j=1}^{J} e^{X_{ij}\beta}}.$$

The log-likelihood function for the individual is:

$$l_i = ln\left(\prod_{j=1}^{J}[Pr_i(j)]^{y_{ij}}\right) = \sum_{j=1}^{J} y_{ij}\, ln[Pr_i(j)]$$

When we have the choice sets for all recreationists, we can sum their log-likelihood functions and apply maximum likelihood to get the estimated coefficients.

The estimated welfare change in RUMs for individual i is:

$$\widehat{\Delta W_i} = \frac{1}{\widehat{\beta_y}}\left\{ln\left[\sum_{\mathrm{j=1}}^{10}\exp\left(X_{ij}{}^{1}\hat{\beta}\right)\right] - ln\left[\sum_{\mathrm{j=1}}^{10}\exp\left(X_{ij}{}^{0}\hat{\beta}\right)\right]\right\}$$

Where $X_{ij}{}^{1}$ and $X_{ij}{}^{0}$ represent the new status and the initial status respectively. Often they are quality changes on one or several sites. $\widehat{\beta_y}$ is the estimated coefficient of income variable, the monetary measure of utility; it equals the negative of the estimated coefficient of travel cost. When a particular site characteristic *l* is changed by one unit at all sites, the welfare measure reduces to $\widehat{\beta_l}/\widehat{\beta_y}$

2.2 Basic Simulation

To simplify the simulations, we assume the recreational sites are beaches and there are three explanatory variables: travel cost (D), beach length (BL) which represents exogenous site characteristics, and facilities (F) which will serve as our potentially endogenous site characteristic. Then the latent utility equation becomes:

$$U_{ij} = D_{ij}\beta_1 + BL_j\beta_2 + F_j\beta_3 + \varepsilon_{ij}$$

Following the estimates reported in Parson (2003), we set "true" values for the population parameters as follows:

$$\beta_1 = -0.06, \beta_2 = 0.49, \beta_3 = 0.06$$

Then the utility equation becomes:

$$(1) \quad U_{ij} = D_{ij} \times (-0.06) + BL_j \times 0.49 + F_j \times 0.06 + \varepsilon_{ij}$$

We assume there are 1,000 people and 10 sites, and the steps of the basic simulation are as follows:

Step I: Take 10,000 random draws for D_{ij} uniformly over the range from 0 to 100, since travel costs are varying across people and sites. Take 10 uniform random draws for BL_j from 0 to 2, and 10 uniform random draws for F_j from 0 to 5, both of which just vary across sites and are the same for all people. These random draws form the pseudo data set for explanatory variables.

Step II: For individual i, extract his/her D_{ij}, BL_j and $F_j, j = 1,2,\ldots,10$, and produce 10 random draws for ε_{ij} from a Type I extreme value distribution with scale factor equal to one. Following Train (2003), the cumulative distribution function for ε_{ij} is:

$$F(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij}))$$

Then its inverse function is: $$\varepsilon_{ij} = -ln(-ln[F(\varepsilon_{ij})])$$

Since $F(\varepsilon_{ij})$ falls between 0 and 1, we can take 10 random draws from a (0, 1) uniform distribution first and then use the inverse CDF function to compute 10 correspondent random numbers for ε_{ij}.

Step III: Use (1) to calculate $U_{ij}, j = 1,2,\ldots,10$. Pick the maximum, mark it as one and others as zero, and we get the pseudo choice variable for individual i.

Step IV: Repeat Step II and III for 1,000 people to obtain the pseudo choice sets and choices for all recreationists, which compose one random sample.

Step V: Regress the pseudo choice variable on the pseudo choice set data set for 1,000 people and get $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$. Do hypothesis tests, where the hypotheses are that the estimated coefficients are equal to their "true" values, and get three t statistics for the three estimates.

Step VI: Repeat Step II, III, IV and V 1,000 times to generate 1,000 random samples, where the explanatory variables remain the same but the error terms are newly drawn for each sample.

Step VII: For the t statistics from 1,000 random samples, calculate the fraction at which they are greater than 1.96, which is the critical value for t statistics at 5% significance level. For the estimated coefficients from 1,000 random samples, calculate the descriptive statistics, such as mean, variance and minimum squared error (MSE).

Table-1 shows the process of simulating individual i's choice set in one random sample.

Table-1: Simulating individual i's choice set

Site	1	2	3	4	5	6	7	8	9	10
D	7.79	61.90	4.23	79.48	56.79	31.95	2.87	71.57	89.71	50.87
F	0.98	4.34	3.48	4.62	2.48	4.98	0.76	1.42	2.45	4.20
BL	1.02	0.64	1.86	1.45	0.90	1.71	0.33	1.59	1.94	1.31
ε	-0.12	0.54	3.61	0.17	7.25	0.62	1.02	0.23	-0.81	1.55
U	-0.03	-2.60	4.48	-3.61	4.43	-0.16	1.05	-3.20	-5.10	-0.61
y	0	0	1	0	0	0	0	0	0	0

According to Cameron and Trivedi (2005), usually there are two types of simulations, one fixed trials and the other with random regressors. The simulation above is the former, but we also try the latter. The steps are very similar, only with a modification to step VI in which we will also repeat step I for each sample. Now, not only the error terms but also the explanatory variables are different for every random sample.

Table-2 and Table-3 show the simulation results for fixed trials and random regressors.

Table-2: Basic simulation results-Fraction of t statistics above 1.96

	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Fixed Trials	0.039	0.053	0.047
Random Regressors	0.047	0.047	0.048

If the null hypotheses are true, each t statistic will follow a standard normal distribution and the fraction at which it is greater than 1.96 should be around 0.05 with a large sample. Here we have 1,000 t statistics for each β. The fractions in Table-2 are all around 0.05, so the Monte Carlo simulations with exogenous explanatory variables generate consistent estimates, both with fixed trials and with random regressors.

Table-3: Basic simulation results-Descriptive statistics

	Fixed Trials			Random Regressors		
	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
True Value	-0.060	0.490	0.060	-0.060	0.490	0.060
Mean	-0.060	0.488	0.060	-0.060	0.490	0.061
Min.	-0.066	0.292	-0.019	-0.067	0.198	-0.094
Max.	-0.054	0.784	0.142	-0.054	0.835	0.173
Var.	4.397e-06	3.965e-03	6.770e-04	4.419e-06	5.638e-03	9.542e-04
MSE.	4.406e-06	3.966e-03	6.764e-04	4.417e-06	5.633e-03	9.545e-04

Also, since $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$ are actually random variables, we can use descriptive statistics to study their properties. In the simulations of fixed trials, their means are very close to their true values (see Table 3). The variance and MSE of $\widehat{\beta_1}$ are almost zero, implying the high precision of simulations on this parameter. $\widehat{\beta_2}$ and $\widehat{\beta_3}$ have wider ranges, likely because these site variables do not vary across individuals resulting in less variation in the data set when compared to the travel cost variable. However, their variances and MSEs are still very small. The descriptive statistics convey the same information as the fractions of t statistics do. In the simulations of random regressors, the results are similar.

3 Simulations with Endogeneity

Next, we make facilities economically endogenous and see whether the estimates still converge to their "true" values. The way we introduce economic endogeneity is to assume there are no facilities at the sites and then the managers will decide the facility levels at each site based on past visitation. Two cases are considered: first, called Case I, when people don't care about facilities, we examine whether the economically endogenous facilities would spuriously affect people's choices (that is, will the estimated conditional logit models suggest a significant parameter estimate for the facilities variable even though the true parameter is zero); second, called Case II, when people do care about facilities, we examine whether the economic endogeneity causes bias in the estimated coefficients.

In Case I, the process of simulations with fixed trials would be different from the one stated in Section 2, and we just list the differences below.

Step I 3a: No data for facilities are created, since there is no facility at the beginning.

Step III 3a: The utility equation used in this step becomes:

$$(2) \quad U_{ij} = D_{ij} \times (-0.06) + BL_j \times 0.49 + \varepsilon_{ij}$$

Step V 3a: This step includes several sub-steps.

1) Average the pseudo choice sets across 1,000 people and get the averaged visit for site j, j=1,2,…,10, denoted by AFP_j
2) Suppose the manager's supply is linearly related with past visitation, and we assume the supply function is:

$$(3) \quad F_j = AFP_j \times 25 + e_j$$

   Since only the relative magnitude of utility matters, we don't include an intercept. 25 is a randomly picked constant. It can be any number. We just want to make sure the scale of newly provided facilities is similar to that of the exogenous facilities in the basic simulation. The error term for the facilities supply function is assumed to have a standard normal distribution, incorporating other factors that may affect facility supply. Take 10 random draws from the standard normal distribution and calculate the facility level using (3) for each site, which is obviously endogenous.

3) This is similar to Step V in part 2. We add the supplied facilities to the pseudo data set, and the true value for β_3 is zero, rather than 0.06.

The rest of the steps are the same. With random regressors, we just need to repeat Step I for each random sample. The results are shown in Table-4 and Table-5.

The fractions of t statistics for $\widehat{\beta_1}$ and $\widehat{\beta_2}$ are around 0.05, so the two estimates are consistent. The fractions for $\widehat{\beta_3}$ are a little bit higher, around 0.08. We could reject the consistency of $\widehat{\beta_3}$ at 5% significance level; however, this is a marginal change in the performance of the conditional logit. We cannot reject that the parameters for facilities are zero if we set the test size as 10%. If we conducted a survey 1,000 times and obtained 1,000 data sets, we would not have a statistically significant impact on the facilities parameter in more than 900 data sets. So, although the endogeneity of facilities does have some effects on $\widehat{\beta_3}$, they are not very substantial.

Table-4: Case 1: Simulation results-Fraction of t statistics above 1.96

	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Fixed Trials	0.043	0.055	0.080
Random Regressors	0.048	0.046	0.084

Table-5: Case 1: Simulation results-Descriptive statistics

	Fixed Trials			Random Regressors		
	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
True Value	-0.060	0.490	0	-0.060	0.490	0
Mean	-0.060	0.469	0.021	-0.060	0.471	0.023
Min.	-0.066	0.163	-0.203	-0.067	0.132	-0.165
Max.	-0.054	0.860	0.165	-0.054	0.839	0.201
Var.	4.302e-06	5.359e-03	2.036e-03	4.314e-06	6.892e-03	1.979e-03
MSE.	4.301e-06	5.801e-03	3.536e-03	4.314e-06	7.254e-03	3.342e-03

With fixed trials, the mean of $\widehat{\beta_1}$ is almost equal to the true value; the variance and MSE are close to zero. The mean of $\widehat{\beta_2}$ is slightly smaller than the true value, and the variance and MSE are bigger than those in the basic simulation. The mean of $\widehat{\beta_3}$, which is the estimated coefficient of the economically endogenous facilities, is greater than the true value as we might expect (though the size of this effect is small). The variance and MSE get bigger, too. $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$ all seem to converge to their true values. So the economic endogeneity does not seem to have much influence in this case. Simulations with random regressors generate similar results. For the types of simulations performed here, facilities do not spuriously affect people's behaviors when the facilities in fact do not matter to the individuals.

In Case II, we now assume that people do care about facilities, so after the facilities are provided, people will update their choice of the best site within their choice sets. We need to account for this in the process of simulations with both fixed trials and random regressors by making the following modifications to the simulation steps:

Step V 3b: After the calculation of endogenous facilities, we add them to the pseudo data set and repeat Step III and IV to get the updated pseudo choice sets for 1,000 people, where the error terms are kept the same and the true β_3 is 0.06. Then the updated pseudo choice sets are used to get estimated coefficients and t statistics.

The results are found in Table-6 and Table-7.

Table-6: Case 2: Simulation results-Fraction of t statistics greater than 1.96

	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Fixed Trials	0.039	0.060	0.078
Random Regressors	0.035	0.045	0.089

Table-7: Case 2: Simulation results-Descriptive statistics

	Fixed Trials			Random Regressors		
	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
True Value	-0.060	0.490	0.060	-0.060	0.490	0.060
Mean	-0.060	0.470	0.080	-0.060	0.470	0.083
Min.	-0.066	0.187	-0.146	-0.067	0.161	-0.112
Max.	-0.053	0.864	0.272	-0.054	0.770	0.252
Var.	4.253e-06	5.470e-03	2.063e-03	4.259e-06	6.761e-03	1.970e-03
MSE.	4.257e-06	5.851e-03	2.455e-03	4.262e-06	7.129e-03	2.470e-03

Case II seems to give the same results. The fraction of t statistics above 1.96 for $\widehat{\beta_1}$ is smaller than 0.05, so we can accept that $\widehat{\beta_1}$ is a consistent estimator of β_1. In the descriptive statistics, the mean of $\widehat{\beta_1}$ is more or less the same as the true value and it has a relatively small range, variance and MSE. Recall that the travel cost variable varies across sites and across individuals which contributes to the robustness of its estimated parameter. The means of $\widehat{\beta_2}$ and $\widehat{\beta_3}$ with both simulations are greater than the true values, coincidentally by around 0.02 for both of them. For $\widehat{\beta_2}$, we can view the differences as slight deviations since the true value of $\beta_2$is 0.49, and $0.02/0.49 \approx 4\%$. The consistency of $\widehat{\beta_2}$ is not affected very much. For $\widehat{\beta_3}$, the much more substantial since the true value of β_3 is 0.06, and $0.02/0.06 \approx 33.33\%$, so the endogeneity has some influence over $\widehat{\beta_3}$, inflating its value by 33%. Since the estimated travel cost parameters are essentially the true values and are estimated very precisely, we would expect that the error in any welfare measures on the endogenously supplied facilities to be driven by the error in the facilities parameter. Despite the 33% increase in the average facilities parameter, the fractions of t statistics above 1.96 for $\widehat{\beta_3}$ is less than 0.10 implying that the chance is more than 90% that we get consistent estimates.

4 Sensitivity Analyses

To investigate how underlying factors in Monte Carlo simulations would influence the simulation results, we conduct sensitivity analysis by changing three elements of the simulation. First, we change the number of sites from 10 to 5 and to 15. Second, we use discrete facilities instead of continuous ones. , Third, we randomly pick numbers as the “true” population parameters rather than use the values from the Parson (2003) study. Since the fractions of t statistics could tell whether the estimates are consistent or not, we just list the fractions here. “RP” stands for randomly drawn parameters, and we pick 5 groups of randomly drawn parameters as the true values for βs with both fixed trials and random regressors. For each group of randomly drawn parameters, β_1 is uniformly drawn over the range of -0.1 and 0; β_2 is uniformly drawn over the range of 0 and 1; β_3 is uniformly drawn over the range of 0 and 0.1. The ranges are chosen with respect to their true values in basic simulations, allowing variations to some extent.

Given an overall review of these data, changing underlying factors of simulations does not change the results very much. The number of sites matter to some extent. As the number of sites grows bigger, the effects of endogeneity become more significant. Variations in the true parameters may have some influence, but if we average across the five groups of randomly assigned true values, the influence may fades. Overall, the patterns observed in the above simulations appear robust for the types of sensitivity analyses conducted here.

Table-8: Sensitivity Analysis-Fractions of t statistics

		Fixed Trials			Random Regressors		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
	5 Sites	0.044	0.049	0.057	0.049	0.045	0.048
	15 Sites	0.051	0.042	0.049	0.047	0.036	0.058
	Discrete	0.040	0.050	0.040	0.055	0.062	0.048
Basic Simulation	*RP 1*	0.049	0.051	0.044	0.043	0.045	0.039
	RP 2	0.041	0.055	0.059	0.044	0.051	0.068
	RP 3	0.054	0.052	0.054	0.055	0.043	0.055
	RP 4	0.036	0.042	0.045	0.034	0.053	0.058
	RP 5	0.054	0.042	0.053	0.058	0.043	0.049
	5 Sites	0.042	0.048	0.057	0.045	0.052	0.054
	15 Sites	0.061	0.058	0.111	0.046	0.050	0.101
	Discrete	0.052	0.048	0.081	0.058	0.052	0.086
Case I Simulation	*RP 1*	0.059	0.053	0.107	0.039	0.057	0.094
	RP 2	0.041	0.066	0.068	0.044	0.077	0.092
	RP 3	0.048	0.078	0.101	0.055	0.051	0.087
	RP 4	0.037	0.059	0.093	0.036	0.074	0.092
	RP 5	0.054	0.045	0.114	0.057	0.050	0.075
	5 Sites	0.048	0.048	0.056	0.048	0.044	0.057
	15 Sites	0.051	0.051	0.090	0.044	0.052	0.092
	Discrete	0.050	0.050	0.070	0.059	0.063	0.077
Case II Simulation	*RP 1*	0.048	0.050	0.099	0.044	0.056	0.087
	RP 2	0.044	0.061	0.072	0.033	0.076	0.078
	RP 3	0.059	0.078	0.093	0.061	0.053	0.066
	RP 4	0.032	0.055	0.086	0.036	0.073	0.097
	RP 5	0.056	0.049	0.106	0.057	0.050	0.075

In the table, almost all the fractions of t statistics of $\widehat{\beta_1}$ are around 0.05. No matter whether facilities are economically endogenous or not, $\widehat{\beta_1}$ is consistent. If we are only interested in the estimated coefficient of travel cost, we may not need to worry about any economic endogeneity in facilities. The fractions of t statistics of $\widehat{\beta_2}$ are a little bit higher than 0.05 in Case I and Case II; the fractions of t statistics of $\widehat{\beta_3}$ are much greater, but smaller than 0.10 most of the time. Thus, as discussed before, the economic endogeneity of facilities would have effects on both estimates of beach length and facilities, which may be attributed to the fact that they both vary across sites but not people. Although the endogenous facilities in Case II can have a substantially inflated effect on the facilities parameter, based on the statistical tests the estimates are generally inconsistent with a probability less than 0.10.

5 Tests for Aggregation Effects

In the econometric sense, the interplay of supply and demand would seem likely to cause simultaneous endogeneity among equations; that is to say, the economic endogeneity of the way that facilities are supplied might be expected to result in statistical endogeneity. However, the results of the simulations do not show strong evidence that the estimates are inconsistent. We notice that the way we made facilities economically endogenous was to build facilities on the average visits of 1,000 people. In other words, facilities are economically endogenous at an aggregate level, and the effect could be diminished when we come to the individual level. Put differently, we assigned the best facilities to the sites that had the highest visitation (i.e., the sites that were best on average). However, the site that is best on average will not best in each individuals choice set, especially given the way we randomly constructed the travel costs. Perhaps the aggregation across all people results in a supply of facilities that remains relatively uncorrelated with what is best in the individual choice sets. To test whether the aggregation across all people influences the results, we change the supply mechanism a little bit. Instead of averaging across all people in one sample, we divide 1,000 people into 10 groups and 100 groups respectively. Under each division principle, we average past visitation within every group, and the facilities are correlated with the group's average visits to each site. Now the economically endogenous facilities are different for different groups. We apply the new mechanism to Case I and Case II. The results are shown in Table-9 and Table-10.

Table-9: Simulation results with endogeneity in different aggregate levels-Fractions of t statistics

		Fixed Trials			Random Regressors		
	Average Across	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Case I	100	0.056	0.444	1	0.053	0.554	1
	10	0.059	1	1	0.062	0.991	1
Case II	100	0.055	0.396	1	0.047	0.504	0.998
	10	0.08	1	1	0.089	0.989	1

Table-10: Simulation results with endogeneity in different aggregate levels-Descriptive Statistics

100 people		Fixed Trials			Random Regressors		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Case I	*True Value*	-0.060	0.490	0	-0.060	0.490	0
	Mean	-0.060	0.361	0.152	-0.060	0.340	0.152
	Min.	-0.067	0.184	0.075	-0.067	0.108	0.071
	Max.	-0.054	0.613	0.240	-0.054	0.590	0.242
	Var.	4.881e-06	3.829e-03	6.303e-04	4.276e-06	3.898e-03	6.329e-04
	MSE.	4.876e-06	2.041e-02	9.086e-03	4.277e-06	2.648e-02	9.070e-03
Case II	*True Value*	-0.060	0.490	0.060	-0.060	0.490	0.060
	Mean	-0.060	0.367	0.201	-0.060	0.345	0.201
	Min.	-0.067	0.193	0.123	-0.067	0.102	0.108
	Max.	-0.054	0.595	0.298	-0.054	0.593	0.296
	Var.	4.997e-06	3.985e-03	6.844e-04	4.230e-06	4.108e-03	6.865e-04
	MSE.	5.007e-06	1.905e-02	2.041e-02	4.256e-06	2.511e-02	2.056e-02

10 people		Fixed Trials			Random Regressors		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
Case I	*True Value*	-0.060	0.490	0	-0.060	0.490	0
	Mean	-0.061	0.155	0.324	-0.061	0.155	0.324
	Min.	-0.068	-0.099	0.284	-0.068	-0.099	0.284
	Max.	-0.053	0.364	0.363	-0.053	0.364	0.363
	Var.	4.961e-06	3.021e-03	1.593e-04	4.961e-06	3.020e-03	1.593e-04
	MSE.	5.791e-06	0.115	0.070	5.791e-06	0.115	0.070
Case II	*True Value*	-0.060	0.490	0.060	-0.060	0.490	0.060
	Mean	-0.062	0.169	0.346	-0.062	0.169	0.346
	Min.	-0.070	-0.066	0.295	-0.070	-0.066	0.296
	Max.	-0.054	0.390	0.387	-0.054	0.390	0.387
	Var.	5.171e-06	3.360e-03	1.732e-04	5.171e-06	3.360e-03	1.732e-04
	MSE.	7.578e-06	0.107	0.082	7.578e-06	0.107	0.082

As we gradually reduce the aggregation level, making the facilties more correlated with the individual choice sets, the effect of economic endogeneity becomes more and more significant. When we average across 100 people, the fractions of t statistics for $\widehat{\beta_3}$ that are above 1.96 are 1, so $\widehat{\beta_3}$ does not converge to the true value and we can reject the null hypotheses in all cases. The fractions of t statistics of $\widehat{\beta_2}$ also increase to around 0.50, so $\widehat{\beta_2}$ is not consistent, either. $\widehat{\beta_1}$, being based on a variable with individual and site specific variation, is not influenced under this situation.

When we average across 10 people, not only are $\widehat{\beta_3}$ and $\widehat{\beta_2}$ inconsistent but $\widehat{\beta_1}$ is also affected some. The fractions of its t statistics go up to 0.08 or 0.09. Although the probability at which $\widehat{\beta_1}$ is not consistent is still small, the endogeneity of facilities does have some spillover effects on the estimated coefficient of travel cost.

We could see how the properties of the estimated coefficients change compared to previous simulations more clearly through the descriptive statistics. For $\widehat{\beta_1}$, when we average across 100 people, the means remain the same as the true value. The variances and MSEs, although still very small, are bigger than previous ones. When we average across 10 people, the means are slightly smaller than the true value, and the ranges keep getting bigger. For $\widehat{\beta_2}$, when we average across 100 people, the means are smaller than the true value; when we average across 10 people, the means become much smaller. The variances and MSEs are getting bigger. The bias is downward. For $\widehat{\beta_3}$, the bias is upward. The means are much greater than the true values. As the aggregation effect declines, the means almost double, with a great increase in MSEs. Thus, when the economic endogeneity of facilities approaches the individual level, the coefficient of beach length tends to be underestimated (attenuated) and the coefficient of facilities tends to be overestimated. Both of the estimates are inconsistent. Actually, since the mean of $\widehat{\beta_1}$ starts to decline when we average across 10 people, we might even expect that $\widehat{\beta_1}$ would become inconsistent with downward bias if we had many more people and the aggregation level was very low.

When the economic endogeneity effect dominates the aggregation effect, the economic endogeneity of facilities will lead to statistical endogeneity, which makes the estimated coefficient of facilities inconsistent. Here, the other two estimated coefficients are influenced, too. It may arise from the fact that facilities are not only correlated with error terms, but also correlated with the other two explanatory variables. A popular site would have a longer beach or be closer to people's houses, and the popularity is proportional to facilities. Thus, facilities are positively correlated with beach length, and negatively correlated with travel cost. So $\widehat{\beta_2}$ would have a downward bias when $\widehat{\beta_3}$ is upward biased; it is also the case with $\widehat{\beta_1}$, since $\widehat{\beta_1}$ is negative. It could be possible that, with a low aggregation level, even if facilities are not included as one explanatory variable in the estimation, as long as they do contribute to people's choices, their economic endogeneity might still have significant effects through their correlations with other site characteristics and travel cost. Plus, within the settings, the estimated marginal welfare change due to a change in facilities at all sites is $-\widehat{\beta_3}/\widehat{\beta_1}$. So if $\widehat{\beta_3}$ is not a consistent estimate, the welfare estimate is also biased, which has important policy implications.

6 Conclusions and Future Study

Site characteristics make contributions to explaining the popularity of recreational sites. Facilities, like parking lots, restrooms, picnic tables and so on, have been identified by previous studies on recreational demand as having a statistically significant on people's utility equations. On the other hand, previous studies on recreational management also show that better facilities are provided at sites where more people go, which means that facilities are typically economically endogenous.

Usually, the interplay of supply and demand will cause simultaneous endogeneity and then lead to inconsistent estimates; however, the Monte Carlo simulations examined here do not strongly support that facilities are statistically endogenous when the supply is based on aggregate demand. In fact, the estimates still converge to the population parameters at a probability of more than 90% even though the mean facilities parameters were overstated by 33%. Because in our simulation design the individuals experience a wide range of travel costs, there likely remain sufficient differences between the sites that are best for an individual and the sites that were best in aggregate. This effect likely minimizes any widespread inconsistency of the facilities parameters even in the endogenous supply case. We test this suspicion by diminishing the aggregation level. The simulation results then become very susceptible to statistical endogeneity of facilities. Therefore, the economic endogeneity effect on the estimation is greatly reduced by the aggregation effect for our simulation.

To clarify, we caution readers against drawing too much from our Case I and Case II simulations results that indicate a high level of consistency of the parameters since the offsetting effect of the aggregation could be caused by the basic property of our simulations. Here we randomly draw numbers as the travel cost, which means that both people and sites are fully dispersed across our hypothetical landscape. That is, our simulations do not involve any spatial clustering of individuals which implies maximal variation in the individual specific travel costs. As a result, on average, the probability of visitation should be almost the same for all sites. And it is the case in our simulation results. When we average the visits across all people, we find that each of the 10 sites has a probability of being visited of about 0.10. So it does not make much difference from the case in which managers do not consider past visitation and construct similar facilities on all sites. On the contrary, it would be common that recreationists cluster at some areas, like cities, and sites such as beaches are dispersed along a shoreline. Then there would be some sites that are more frequently visited than others. In fact, this situation is much closer to reality. Lupi and Feather (1998) put recreational sites into three categories: the most popular ones, the ones which are the subject of policy analysis and remaining sites. In their survey of sport fishing in Minnesota, Lake Mille Lacs dominated all other lakes; Lake of the Woods, Lake Minnetonka and Lake Leech were the second popular; when it came to other lakes, the number of visitors dramatically declined. So, in this case, even if we aggregate across all people, the popular sites wouldn't disappear.

When dealing with real data, if there is a large degree of spatial dispersion among recreationists when compared to site locations, we might be inclined to neglect the possible economic endogeneity of facilities; however, if the dispersion of recreationists is limited (perhaps because they live in spatially clustered regions so will have similar travel costs) then the economic endogeneity may be more problematic. In such situations, statistical techniques such as apply instrumental variables may be warranted. Future directions for our subsequent investigations include incorporating these situations of disproportionately popular sites and spatial clustering of travel costs into our simulations and investigating whether the economic endogeneity of facilities will cause statistical endogeneity under those circumstances. That examination will provide more robust conclusions which would allow researchers to identify the types of situations likely to cause more or less of a concern about the impacts of endogenously supplied site characteristics.

References

Allen, P. G., T. H. Stevens, et al. (1981). “The effects of variable omission in the travel cost technique.” Land Economics 57(2): 173-180.

Ben-Akiva, M., and S. Lerman. 1985. Discrete choice analysis: Theory and application to travel demand, MIT Press: Cambridge, MA.

Brown, W. G. and F. Nawas (1973). “Impact of aggregation on the estimation of outdoor recreation demand functions.” American Journal of Agricultural Economics 55(2): 246-249.

Cameron A.C., P.K. Trivedi, Chapter 7 “Hypothesis tests.”, Microeconometrics: Methods and Applications, Cambridge University Press 2005: 250-254

Caner, M. and B. E. Hansen (2004). “Instrumental variable estimation of a threshold model.” Econometric Theory 20(5): 813-843.

Caulkins, P. P., R. C. Bishop, et al. (1985). “Omitted cross-price variable biases in the linear travel cost model: correcting common misperceptions.” Land Economics 61(2): 182-187.

Cesario, F. J. and J. L. Knetsch (1970). “Time bias in recreation benefit estimates.” Water Resources Research 6(3): 700-&.

Cook, A. (2008). “Recreation value of a new long-distance walking track.” Tourism Economics 14(2): 377-391.

Cutter, W. B., L. Pendleton, et al. (2007). “Activities in models of recreational demand.” Land Economics 83: 370-381.

Faghri, A., A. Lang, et al. (2002). “Integrated knowledge-based geographic information system for determining optimal location of park-and-ride facilities.” Journal of Urban Planning and Development-Asce 128(1): 18-41.

Gum, R. L. and W. E. Martin (1975). “Problems and solutions in estimating the demand for and value of rural outdoor recreation.” American Journal of Agricultural Economics 57(4): 558-566.

Haab, T. C. and R. L. Hicks (1997). “Accounting for choice set endogeneity in random utility models of recreation demand.” Journal of Environmental Economics and Management 34(2): 127-147.

Hellerstein, D. (2005). “Modeling discrete choice with uncertain data: an augmented MNL estimator.” American Journal of Agricultural Economics 87(1).

Hicks, R. L. and K. E. Schnier (2006). “Dynamic random utility modeling: a monte carlo analysis.” American Journal of Agricultural Economics 88(4): 816-835.

Lew, D. K. and D. M. Larson (2005). “Accounting for stochastic shadow values of time in discrete-choice recreation demand models.” Journal of Environmental Economics and Management 50(2): 341-361.

Lupi, F. and P. M. Feather (1998). "Using partial site aggregation to reduce bias in random utility travel cost models." Water Resources Research 34(12): 3595-3603.

Martha E. Lee, B. L. Driver, Chapter 10 “Benefits-based management: a new paradigm for managing amenity resources.”, Ecosystem Management: Adaptive Strategies for Natural Resources Organizations in the Twenty-first Century, Taylor & Francis, 1999: 143-154.

McConnell, K. E. and V. A. Duff (1976). “Estimating net benefits of recreation under conditions of excess demand.” Journal of Environmental Economics and Management 2(3): 224-230.

Murdock, J. (2006). “Handling unobserved site characteristics in random utility models of recreation demand.” Journal of Environmental Economics and Management 51(1): 1-25.

Shin, W. S., R. Jaakson, et al. (2001). “Benefits-based analysis of visitor use of Sorak-San National Park in Korea.” Environmental Management 28(3): 413-419.

Timothy C. Haab, Kenneth E. McConnell, Chapter 8 “Site choice models.”, Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation, Edward Elgar Publishing 2002: 209-212.

Timmins, C. and J. Murdock (2007). “A revealed preference approach to the measurement of congestion in travel cost models.” Journal of Environmental Economics and Management 53(2): 230-249.

Train Kenneth, Chapter 9 “Drawing from densities.”, Discrete Choice Methods with Simulation, Cambridge University Press 2003: 209-210.

von Haefen, R. H. (2008). “Latent consideration sets and continuous demand systems.” Environmental & Resource Economics 41(3): 363-379.

von Haefen, R. H. and D. J. Phaneuf (2008). “Identifying demand parameters in the presence of unobservables: a combined revealed and stated preference approach.” Journal of Environmental Economics and Management 56(1): 19-32.

von Haefen, R. H., D. M. Massey, et al. (2005). “Serial nonparticipation in repeated discrete choice models.” American Journal of Agricultural Economics 87(4).

Wetzel, J. N. (1977). “Estimating the benefits of recreation under conditions of congestion.” Journal of Environmental Economics and Management 4(3): 239-246.

Yeh, C.-Y., T. C. Haab, et al. (2006). “Modeling multiple-objective recreation trips with choices over trip duration and alternative sites.” Environmental and Resource Economics 34: 189-209.