DOES ECONOMIC ENDOGENEITY OF SITE FACILITIES IN RECREATION DEMAND MODELS LEAD TO STATISTICAL ENDOGENEITY?

By

Min Chen

A PLAN B PAPER

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Agricultural Economics

2009

ABSTRACT

DOES ECONOMIC ENDOGENEITY OF SITE FACILITIES IN RECREATION DEMAND MODELS LEAD TO STATISTICAL ENDOGENEITY?

By

Min Chen

Random Utility Models of recreation demand are widely used to relate demand and value to the characteristics of recreation sites. Although some kinds of endogeneity problems have been studied in previous literature, no study has addressed the potential problem with site characteristics that are endogenously supplied. Some site characteristics, like facilities, could be endogenous in an economic sense due to the interplay of supply and demand. That is, 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 this economic endogeneity might cause problems for econometric models linking site demand to facilities. In this paper, we use Monte Carlo simulations to investigate under what situations this economic endogeneity will lead to statistical endogeneity.

ACKNOWLEDGEMENTS

First of all, I want to thank my master advisor, Dr. Frank Lupi, for his extremely great advice and support in the completion of this paper.

Second, I appreciate the constructive comments given by my master committee members, Dr. Patricia E. Norris and Dr. Jinhua Zhao.

Also, I thank the Michigan Experiment Station, the Michigan State University Environmental Science and Policy Program, and the NOAA Coastal Stressors program for funding this research.

Finally, I wish to express my thankfulness to my parents and family in China, for their love and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section 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 an 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 focused on the possible types of endogeneity in RUMs. Following Ben-Akiva and Lerman (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 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 a variable for 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 that are supplied by managers in response to people's use of a recreation site, 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.

While there is empirical evidence that facilities can affect estimated demand, the supply and types of facilities are also determined by people's visitation as the literature in park 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 build a walking track since its value was low. All of these demonstrate that facilities are more likely to be built on sites where people go.

This may appear similar to the congestion variable: congestion also happens on popular sites, discouraging future visitation though. But the mechanisms are not the same, and the ways to model the endogeneity of congestion and facilities are different. The fact that congestion depends on visits comes from the externality problem: one person's visit has negative effects on others. The level of congestion on one site is determined by people's behaviors, so it looks more like a game theory context. For facilities, however, a site manager is involved. If we view the managers as the supply side and the recreationists as the demand side, managers change facilities in response to recreation demand, and recreation demand varies in response to facilities. The interplay of supply and demand makes facilities endogenous in the economic sense.

Notice that facilities are site-specific, so they are endogenous at an aggregate level rather than the individual level. In fact, the endogeneity problem of facilities looks like the price endogeneity problem in market demand models, which has received lots of attention in market analysis literatures. Dhar, Chavas and Gould (2003) stated that price endogeneity was an issue in the estimation of aggregate demand functions, resulting in simultaneous equation bias. Yang, Chen and Allenby (2003) mentioned that researchers had found that failure to account for price endogeneity led to estimation bias in both aggregate and disaggregate data. Especially, Villas-Boas and Winter (1999) tested the endogeneity of market-mixed variables set by market managers in random utility models, and showed that even if those variables were common across all consumers, or individuals were price takers and strategically didn't impact the price setting behavior of the sellers, endogeneity would still be a potential problem. Therefore, in recreation demand models, although everyone faces the same facilities on all sites, the economically endogenous facilities may lead to statistical endogeneity. Accordingly the objective of this paper is to examine in what circumstances the economic endogeneity of facilities causes problems to the estimation results.

To address this issue, Monte Carlo simulations are applied. In the simulations, we set values for the “true” parameters, simulate choices, run regressions, and obtain estimates. If there is little bias in the estimates compared with true parameters, the economic endogeneity of facilities does not matter. If the differences between estimates and true parameters are huge, then facilities are not only economically endogenous but also statistical 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 whether an estimate is biased for sure without knowing its true value.

In the following sections, we present the basic choice model for our recreation demand simulations, and state our extrapolation of possible factors having influence on how much economic endogeneity will be reflected on statistical results. For the simulations, we assume all explanatory variables except facilities are exogenous, let facilities be determined by recreation demand and supply, and test whether the estimates are biased under a variety of situations. Also, some underlying attributes of simulations are changed in order to see how sensitive the results are. Finally, we conclude our simulation studies after a discussion of those simulation results.

Section 2

METHODS

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, utility maximization implies 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 indicating the chosen site:

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

According to McFadden (1974), when ε_{ij} follows a Type I extreme value distribution, the maximization of the random utilities yields site choice probabilities that are 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 choices for all recreationists, we can sum their log-likelihood functions and apply maximum likelihood estimation method to get the estimated coefficients.

Welfare Measures

Researchers are often interested in how people value the loss of a certain site or the welfare change of recreationists corresponding to an environmental quality change. For individual i, if we know ε_{ij} , where j = 1,2,…,J, we can calculate both of these welfare measures. Recall that recreation demand models can measure only use values and suppose site k is closed due to some reason; this person puts no value on its loss if site k does not give the highest utility (i.e., if site k is not being used). If site k is individual i's choice, due to its closure, i would then go to the site with the second highest utility, say, site m. Let $X_{ij} = [(M_i - TC_{ij}), x_{ij}]$, where $(M_i - TC_{ij})$ is individual i's income minus travel cost to site j, which is also the expenditure on other commodities, and x_{ij} contains the covariates for site characteristics. The compensating welfare measure can be expressed by:

$$U_{ik}[(M_i - TC_{ik}), x_{ik}] = U_{im}[(M_i - TC_{im} + C_i), x_{im}]$$

Given the utilities are known, we have:

$$\beta_M(M_i - TC_{ik}) + \beta_x x_{ik} + \varepsilon_{ik} = \beta_M(M_i - TC_{im} + C_i) + \beta_x x_{im} + \varepsilon_{im}$$

$$U_{ik} = U_{im} + \beta_M C_i$$

$$C_i = \frac{U_{ik} - U_{im}}{\beta_M}$$

Where β_M is the income parameter, the monetary measure of utility, which is the absolute value of the travel cost parameter. So the access value of one site is equal to the reduction in utility divided by marginal utility of income if it is visited by someone; otherwise, its value is zero. When we average the values across people, we will get the average site value for each site in one sample.

As for the value of environmental quality change, assume there is a change of quality *l* on site k, which is small enough to not to make individual i switch to another site if it is a bad thing. Following the example above, now we have m=k after the change, with $x_{im} = x_{ik}^{'} = x_{ik} + \Delta x_l$. Similarly, we will get:

$$C_{qi} = \frac{U_{ik} - U_{ik}^{'}}{\beta_M} = -\frac{\beta_l}{\beta_M}\Delta x_l$$

If it is bad then the compensating welfare measure is positive, and vice versa. When it is a marginal change, this welfare measure becomes β_l/β_M. Given that we are measuring use values, for site j, where j≠k, there is no value for its quality change for individual i. Hence, we can get the welfare measure vector simply by multiplying the individual choice vector by β_l/β_M. Again, averaging the vectors across people gives the average welfare measures of marginal quality change for all sites.

In empirical studies, however, there is no way to know the individual error terms. We are able to get the estimated values of the welfare measures instead of true values. In general, under the conditional logit model, the estimated welfare change for individual i caused by any change in the covariates is:

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

Where ${X_{ij}}^{1}$ and ${X_{ij}}^{0}$ represent the new status and the initial status respectively. $\widehat{\beta_M}$ is the estimated coefficient of income variable.

The estimated marginal value of environmental quality change for individual i is:

$$\partial\widehat{W_i}/\partial q_j = \frac{\widehat{\beta_l}}{\widehat{\beta_M}}\widehat{Pr_i}(j), j = 1,2, \ldots, J$$

Where $\widehat{Pr_i}(j)$ is the predicted probability of individual i to visit site j. The estimates won’t be zero for any of the sites due to the predicted probability. We can get the average estimates by averaging those individual estimates across people.

<u>**Basic Steps**</u>

To simplify the model, we assume there are three explanatory variables: travel cost (TC), quality (Q) which represents exogenous site characteristics, and facilities (F) which will serve as our potentially endogenous site characteristic. The latent utility equation becomes:

$$U_{ij} = TC_{ij}\beta_1 + Q_j\beta_2 + F_j\beta_3 + \varepsilon_{ij}$$

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

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

Then the utility equation becomes:

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

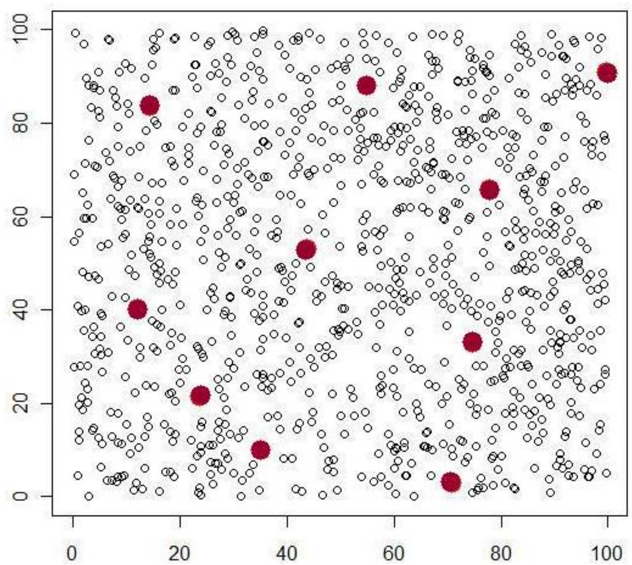

Figure 2-1. Basic Landscape

We also assume there are 10 recreation sites and 1,000 recreationists. Both sites and people are randomly spread out in a certain area. Figure 2-1 illustrates their locations, where the red dots represent sites and small black dots are individuals.

In general, basic steps for the simulation under these settings with exogenous independent variables will be as follows:

Step I: Take 10,000 random draws for TC_{ij} uniformly over the range from 0 to 100, since travel costs are varying across people and sites. Take 10 uniform random draws for Q_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. The ranges for the covariates are set to roughly correspond to the data in Parson (2003).

Step II: For individual i, extract his/her TC_{ij}, Q_j and $F_j, j = 1,2, \dots, 10$, and produce 10 random draws for ε_{ij} from a Type I extreme value distribution with a variance of $\pi^2/6$. 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 $U_{ij}, j = 1,2, \dots, 10$. 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 pseudo choices for all recreationists, which compose one random sample.

Step V: Regress the pseudo choice variable on the pseudo data set for 1,000 people and get $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$. Do hypothesis tests, where the null hypotheses are that the estimated coefficients are equal to their "true" values, and the significance level is chosen to be 5%. With the sample size equal to 1,000, the critical value for t statistics at 5% significance level is 1.96.

Step VI: Repeat Step II, III and IV 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. Do Step V for each random sample, producing 1,000 $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$, which can be viewed as samples for three random variables.

Step VII: For the 1,000 iterations, each with one random sample, calculate the percent of times in which the null hypotheses are rejected at 5% significance level, that is, the t statistics are greater than the critical value 1.96. For estimated coefficients from the 1,000 random samples, calculate the descriptive statistics, such as mean, variance and mean squared error (MSE).

Table 2-1. Simulating Individual i's Choice

Site	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
TC	7.79	61.90	4.23	79.48	56.79	31.95	2.87	71.57	89.71	50.87
Q	1.02	0.64	1.86	1.45	0.90	1.71	0.33	1.59	1.94	1.31
F	0.98	4.34	3.48	4.62	2.48	4.98	0.76	1.42	2.45	4.20
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

Table 2-1 shows the process of simulating individual i's choice in one iteration. Utilities are computed with the pseudo data for travel cost, site quality and facilities and randomly drawn errors. This individual will choose site 3 since it provides the highest utility. Note that travel cost and errors vary across sites and people, but quality and facilities only vary across sites, so they stay the same for different people.

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

What Matters?

Normally, the endogeneity of a variable makes it correlated with the errors and thus the estimates are biased. Here, whether facilities become highly correlated with other explanatory variables also matters. As stated in the introduction, facilities only vary across sites. Their economic endogeneity occurs at an aggregate level, so there might not be a lot of correlations between facilities and errors, or to say, the individual choices may not affect the levels of facilities very much. The endogeneity effect could be wiped out in the regression of individuals. Thus, the fact that the facility variable is site-specific plays an important role.

On the other hand, other site characteristics, like quality, are also the same for all people. Given managers build facilities based on past visitation partly determined by those characteristics, it won't be surprising if we see a big correlation between facilities and those variables. If this happens, even though facilities are not correlated with individual errors, the multicollinearity of regressors may cause problems to estimates, too. Usually, there are many other factors for managers to take into account while investing in facilities, like budget, cost of maintenance, etc. When the economic endogeneity of facilities is stronger, facilities are more likely to depend solely on past visitation, as a result of which the correlation between facilities and other site characteristics will be higher, and then we would see bigger bias in our estimates. Hence, in our simulations, we incorporate different levels of the strength of economic endogeneity through the supply of facilities to investigate whether the correlation between facilities and other site characteristics has substantial influence over the statistical results.

Plus, how important facilities are in the utility equation could also take a part. In the utility equation:

$$U_{ij} = TC_{ij}\beta_1 + Q_j\beta_2 + F_j\beta_3 + \varepsilon_{ij}$$

β_1 is negative while the other two betas are positive. In many applications, it would not be uncommon to find that $|TC_{ij}\beta_1| > Q_j\beta_2$ and $|TC_{ij}\beta_1| > F_j\beta_3$. That is, the travel cost portion of indirect utility is relatively large in magnitude and plays a major role in determining the utility level. But if facilities were relatively more important in utility, say for example, if facilities were scaled to be more important than travel costs, that is, $F_j\beta_3 > |TC_{ij}\beta_1|$, their economic endogeneity might cause more problems.

Section 3

MONTE CARLO SIMULATIONS

Simulations with Endogeneity

Following the basic steps described in Section 2, in our simulations, we make facilities correlated with past visitation. The way we introduce their economic endogeneity is to assume there are no facilities at the sites and then managers determine the facility levels at each site based on past visitation. In order to account for some heterogeneity, two types of worlds are considered. In the first one, called Case I, people don't care about facilities, and 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). In the second one, called Case II, people do care about facilities, and we examine whether the economic endogeneity causes bias in the estimated coefficients. The true parameter of the facility variable is 0 in Case I and 0.06 in Case II.

In Case I, the process of simulations with fixed regressors is somewhat different from the one stated in Section 2. 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:

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

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

1) Average the pseudo choices across 1,000 people and get the averaged visit for site j, j=1,2,…,10, denoted by $\bar{y}_j$:

$$\bar{y}_j = \frac{1}{1000}\sum_{i=1}^{1000} y_{ij} \ , j = 1,2, \dots ,10$$

2) Suppose the manager's supply is linearly related with past visitation, and we assume the supply equation is:

$$(3) \qquad F_j = \bar{y}_j \alpha + e_j$$

Since only the relative magnitude of utility matters, we don't include an intercept in (3). The error term for the facilities supply function is assumed to have a standard normal distribution, incorporating other factors that may affect facility supply, like budget constraint, cost of maintenance, etc.

The coefficient α can be any number, and it is related to the correlation between facilities and average past visitation.

$$Corr(F_j, \bar{y}_j) = \frac{Cov(F_j, \bar{y}_j)}{\sqrt{Var(F_j)Var(\bar{y}_j)}} = \frac{\alpha\sqrt{Var(\bar{y}_j)}}{\sqrt{\alpha^2 Var(\bar{y}_j) + Var(e_j)}}$$

Due to the existence of the error term, the correlation increases as α gets bigger. We pick several values for α from 10 to 200 to test how sensitive the results are to the correlation.

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.

In Case II, we assume that people do care about facilities, so after 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 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 choices for 1,000 people, where the error terms are kept the same and the true β_3 is 0.06. Then the updated pseudo choices are used to get estimated coefficients and t statistics.

For the two types of worlds, the rest of the simulations are the same as basic steps. We also run simulations with random regressors. Since the results are very similar, we don't show them for these and the following simulations, and all discussions will focus on results from simulations with fixed regressors.

From Table 3-1, the economic endogeneity of facilities seems to be more of a problem as the coefficient in the supply equation increases. Intuitively, the larger the correlation between facilities and past visitation is, the stronger the economic endogeneity is, as a result of which the statistical bias results become more substantial.

Look at $\widehat{\beta_1}$ first. In all situations, it has a mean equal to its true value, so there is no bias in the estimated coefficient for the travel cost variable. It will have unbiased standard errors, too, for the variance and MSE are almost the same. The probability to reject the true value is very low, around 5%. $\widehat{\beta_1}$ is not affected by the endogenous

Table 3-1. Simulation Results with Endogeneity

		Case I			*Case II*		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
α=10	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.49	0.01	-0.06	0.49	0.07
	Var.	4.3e-06	3.9e-03	2.1e-03	4.3e-06	3.9e-03	2.1e-03
	MSE.	4.3e-06	3.9e-03	2.3e-03	4.3e-06	3.9e-03	2.3e-03
	*Percent(%)**	3.8	5.7	5.6	3.9	5.3	4.8
α=25	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.47	0.02	-0.06	0.47	0.08
	Var.	4.3e-06	5.4e-03	2.0e-03	4.3e-06	5.5e-03	2.1e-03
	MSE.	4.3e-06	5.8e-03	2.5e-03	4.3e-06	5.9e-03	2.5e-03
	Percent(%)	4.3	5.5	8.0	3.9	6.0	7.8
α=45	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.43	0.04	-0.06	0.43	0.09
	Var.	5.1e-06	0.008	1.4e-03	5.1e-06	0.009	1.4e-03
	MSE.	5.1e-06	0.012	2.8e-03	5.2e-06	0.012	2.6e-03
	Percent(%)	6.2	9.1	15.7	6.5	9.9	13.4
α=70	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.38	0.05	-0.06	0.38	0.10
	Var.	4.2e-06	0.013	1.1e-03	4.2e-06	0.014	1.2e-03
	MSE.	4.2e-06	0.026	3.1e-03	4.2e-06	0.026	3.0e-03
	Percent(%)	4.7	15.9	23.8	4.8	14.3	21.0
α=100	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.29	0.05	-0.06	0.30	0.11
	Var.	4.2e-06	0.015	6.3e-04	4.1e-06	0.016	7.0e-04
	MSE.	4.2e-06	0.055	3.1e-03	4.1e-06	0.051	2.7e-03
	Percent(%)	5.4	27.6	37.4	3.9	24.1	30.9
α=200	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.23	0.03	-0.06	0.26	0.09
	Var.	4.4e-06	0.010	1.3e-04	4.5e-06	0.017	2.0e-04
	MSE.	4.4e-06	0.079	1.2e-03	4.5e-06	0.071	9.6e-04
	Percent(%)	4.5	38.8	49.1	4.5	30.0	34.4

*Percent of times to reject null hypotheses at 5% significance level that the estimates are equal to their true values

facilities. This makes sense, since travel cost varies both across people and sites and it has the biggest variation. In fact, if we set the sample size large enough, the fraction of t statistics can be compared with the level of significance to determine whether the estimate is actually consistent, because when an estimate is consistent, the probability to reject its true value should be close to the chosen significance level.

For $\widehat{\beta_2}$, when α is 10, the mean is equal to its true value and the variance is equal to the MSE. There is no bias in the estimated coefficient for the quality variable and the standard errors. The percents are around 5%. So $\widehat{\beta_2}$ has a good performance in this situation. When α is 25, the percents are still around 5%; however, the bias in the mean starts to come out. The mean is smaller than the true value by 0.02, which is a 4% downward bias. The MSE is a little greater than the variance. When α increases to 45, there is a 12% downward bias in the mean, and the variance and MSE are also growing. The percents are close to 10%. Although this is bigger than 5%, more than 90% of the time we cannot reject the true value of $\widehat{\beta_2}$. As α keeps increasing, the bias in the mean becomes larger. The difference between variance and MSE is also getting bigger, but both of them are in the same magnitude as those with α equal 45. When α goes up, it is more likely to reject the true value of $\widehat{\beta_2}$. The probability is about one third with α equal to 200.

$\widehat{\beta_3}$ is the estimated coefficient for the facility variable, which is hypothesized to be influenced by the endogeneity. When α is 10, the mean is greater than the true value by 0.01. In Case II, that is a 17% upward bias. The variance is smaller than the MSE, so there is also some bias in the standard errors. The probability to reject the true value of $\widehat{\beta_3}$ is around 5%. When α is 25, the bias in the mean in Case II grows to 33%. The variance and MSE do not change much though. The chances to reject the true value of $\widehat{\beta_3}$ are still less than 10%. When α is 45, in Case II, there is a 50% upward bias. The variance and

MSE are lower to some extent. The percents now exceed 10%. When α is 70 or even bigger, the bias of the mean could reach almost 83%, or at least 67%. The variance and MSE are getting smaller as α gets bigger, but the differences between the two are very obvious in these situations. The percents are almost 50% in Case I when α is 200, and close to 40% in Case II. Thus, nearly half of the time we would reject the true value of $\widehat{\beta_3}$.

Based on the results, we can see that as the correlation between facilities and past visitation increases, there is more bias in the estimated coefficients, and the probability to reject their true values is getting bigger, too. When the coefficient in the facility supply equation is small, the economic endogeneity of facilities is not a big problem. When this coefficient is very big, not only the estimate for the facility variable, but also the estimate for the quality variable is affected. Facilities' economic endogeneity has spillover effects. In addition, comparing the results derived in Case I and Case II, they are similar, but it is easier to reject the true values of the estimates in Case I. So if people don't care about facilities and we still put that variable in the estimation, we are more likely to have trouble with economically endogenous facilities than when people do care about facilities.

Correlations

The correlation between facilities and past visitation becomes larger when the coefficient in the facility supply equation increases. As discussed in Section 2, the correlation between facilities and past visitation could lead to multicollinearity with other site-specific variables beside correlation with errors. When the endogeneity of facilities is at the aggregate level, the correlation with errors is small, but as the correlation between facilities and past visitation increases, which is equivalent to saying that the economic endogeneity gets stronger, the correlation with other site characteristics may go up.

Particularly, here an individual's past visitation to each site is negatively correlated with travel cost and positively correlated with quality. Given facilities are built on average past visitation, they should also be correlated with travel cost and quality to some extent, although the correlations may not be in the same magnitude. And we would see larger correlations with a bigger α. To illustrate this point, we compute correlations of endogenous facilities with travel cost and quality.

Table 3-2. Correlations of Facilities with Travel Cost and Quality

	α=10	*α=25*	*α=45*	*α=70*	*α=100*	*α=200*
*TC-F**	0	0	0	0.01	-0.02	-0.01
*Q-F***	0.27	0.57	0.71	0.72	0.86	0.91

*Correlation between travel cost and facilities; **Correlation between quality and facilities

From Table 3-2, we could find that facilities and travel cost are almost independent, no matter how big the supply coefficient is. This is possible, since travel cost has a greater variation than facilities do. There are only 10 levels of facilities, but 1,0000 different travel costs. Plus, in the landscape where both people and sites are randomly located, individual choices are very different. The best site for one person could be the worst for another. When we average them to get average past visitation, the effect of travel cost becomes minimal. Since facilities are built based on the average visitation, the correlation between facilities and travel cost is almost zero.

When it comes to quality, another site characteristic, we see that the endogeneity of facilities can induce a multicollineartiy problem. Since all recreationists face the same quality and prefer a site with better quality, the effect of quality will not be wiped out by averaging. When α is 10, the correlation is small, close to 0.30. When α is 25, the correlation is larger than 0.5, but we do not see much bias of the estimated preferences yet. The results become substantial when α grows to 45, and the correlation between facilities and quality is as high as 0.7. It keeps increasing as α gets bigger, which is what we would expect. So in this situation, the multicollinearity between facilities with quality needs to be very high to cause bias to the estimates. Then in empirical studies, we can use this correlation for the purpose of diagnosis when we have the landscape of random location. If the correlation is too high, we might suspect the facilities to be economically endogenous, and this economic endogeneity would be strong enough to cause problems to the estimation.

Site-Specific Property

In both Section 1 and Section 2, we mentioned that the fact that facilities are endogenous at an aggregate level plays an important role in whether the endogeneity is worth worrying about. To test this, 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. We do the regression for all groups as a whole, so facilities are no longer site-specific, but group-specific. We apply the new mechanism to Case I and Case II under the situation where α is 25.

Table 3-3. Simulation Results with Group-Specific Facilities

		Case I			Case II		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
10 Groups	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.36	0.15	-0.06	0.37	0.20
	Var.	4.9e-06	3.8e-03	6.3e-04	5.0e-06	4.0e-03	6.8e-04
	MSE.	4.9e-06	2.0e-02	0.024	5.0e-06	1.9e-02	0.020
	Percent(%)	5.6	44.4	100	5.5	39.6	100
100 Groups	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.14	0.33	-0.06	0.15	0.35
	Var.	4.9e-06	2.2e-03	1.4e-04	5.0e-06	2.4e-03	1.6e-04
	MSE.	5.6e-06	0.13	0.11	7.1e-06	0.12	0.083
	Percent(%)	5.9	100	100	8.0	100	100

When we divide people into 10 groups, $\widehat{\beta_1}$ remains unaffected with no bias in the mean and standard error. The percents are a little bit greater than 5%. For $\widehat{\beta_2}$, there is a more than 20% downward bias. The variance and MSE are different, and the probability to reject its true value is around 40%. For $\widehat{\beta_3}$, the percents are 100%, so we can reject its true value for sure. In Case II, the upward bias is more than two times of the true value itself. The variance and MSE are not even in the same magnitude.

When we further divide people in 100 groups, even $\widehat{\beta_1}$ gets affected. Although the mean is the same as its true value, there are some differences between the variance and MSE. The percents are still lower than 10%, but in Case II, the percent already grows to 8%, which is quite big for $\widehat{\beta_1}$. In this situation, the probability to reject the true value of $\widehat{\beta_2}$ is 100%. Its downward bias in the mean is larger than 60%. The MSE is at least 100 times of the variance. $\widehat{\beta_3}$ is severely biased with a greater bias.

Table 3-4 shows estimated correlations of facilities with travel cost and quality. Endogenous facilities are not correlated with the two exogenous variables very much, but the economic endogeneity definitely cause bias to at least $\widehat{\beta_2}$ and $\widehat{\beta_3}$.

Table 3-4. Correlations of Facilities with Travel Cost and Quality

	10 Groups	*100 Groups*
TC-F	-0.01	-0.09
Q-F	0.36	0.17

When facilities are supplied on a more individual-specific basis, individual unobserved preferences become more important and endogenous facilities are more highly correlated with the errors than with other variables, which leads to the typical statistical endogeneity problem. That’s why the correlation of facilities with quality decreases as the number of groups increases, and the correlation with travel cost goes up a little bit, because travel cost is individual-specific. We don’t need the coefficient in the facility supply equation to be very big to see a huge bias in the estimates. On the contrary, when facilities are provided at the aggregate level, idiosyncratic effects are averaged out and other site characteristics become the key factors. But the induced multicollinearity has to be very severe to cause problems in the estimation, at least in the landscape with random location. Therefore, this site-specific property of facilities greatly diminishes the effect of the economic endogeneity.

Section 4

SENSITIVITY ANALYSES

To investigate how other basic settings in Monte Carlo simulations would influence the simulation results, we conduct sensitivity analyses by changing some elements of the simulation. For example, we change the number of sites from 10 to 5 and to 15. We use discrete facilities instead of continuous ones. Also, we randomly pick numbers as the "true" population parameters rather than use the values from the Parson (2003) study. We pick several groups of randomly drawn parameters as the true values for βs. 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 previous simulations, allowing variations to some extent. Simulations are done under situations when α equals 25 and 70. In the statistics below, we just show the means and percent of times we reject the null hypotheses that estimated parameters equal their true values.

With a general view of these data, changing these settings of simulations does not change the results very much. When we assume the facilities to be discrete, or we change the true values of the parameters, the results are similar. The mean of $\widehat{\beta_2}$ has a downward bias and the mean of $\widehat{\beta_3}$ has an upward bias. There is no bias in the mean of $\widehat{\beta_1}$. Chances to reject the true value of $\widehat{\beta_3}$ are greater than those to reject the true value of $\widehat{\beta_2}$. The results in Case I seem to be more substantial than in Case II.

Table 4-1. Results of Sensitivity Analyses

			Case I			*Case II*		
			$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
5 Sites	*α=25*	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.45	0.03	-0.06	0.45	0.08
		Percent(%)	4.2	4.8	5.7	4.8	4.8	5.6
	α=70	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.34	0.03	-0.06	0.36	0.09
		Percent(%)	5.7	10.6	10.5	5.8	9.8	10.4
15 Sites	*α=25*	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.48	0.03	-0.06	0.48	0.08
		Percent(%)	6.1	5.8	11.1	5.1	5.1	9.0
	α=70	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.38	0.05	-0.06	0.39	0.11
		Percent(%)	5.6	15.8	37.7	5.2	15.8	30.6
Discrete Facilities	*α=25*	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.47	0.02	-0.06	0.48	0.08
		Percent(%)	5.2	4.8	8.1	5.0	5.0	7.0
	α=70	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
		Mean	-0.06	0.38	0.05	-0.06	0.39	0.10
		Percent(%)	3.7	16.6	27.8	3.8	15.2	21.8
Random Parameter 1	*α=25*	*True Value*	-0.01	0.24	0	-0.01	0.24	0.08
		Mean	-0.01	0.22	0.02	-0.01	0.23	0.10
		Percent(%)	5.9	5.3	10.7	4.8	5.0	9.9
	α=70	*True Value*	-0.01	0.24	0	-0.01	0.24	0.08
		Mean	-0.01	0.16	0.05	-0.01	0.16	0.12
		Percent(%)	4.2	10.0	30.3	5.6	8.4	27.8
Random Parameter 2	*α=25*	*True Value*	-0.04	0.95	0	-0.04	0.95	0.04
		Mean	-0.04	0.90	0.02	-0.04	0.90	0.06
		Percent(%)	4.1	6.6	6.8	4.4	6.1	7.2
	α=70	*True Value*	-0.04	0.95	0	-0.04	0.95	0.04
		Mean	-0.04	0.67	0.05	-0.04	0.68	0.09
		Percent(%)	4.1	24.1	29.2	3.9	22.5	25.9
Random Parameter 3	*α=25*	*True Value*	-0.03	0.53	0	-0.03	0.53	0.06
		Mean	-0.03	0.50	0.03	-0.03	0.50	0.08
		Percent(%)	4.8	7.8	10.1	5.9	7.8	9.3
	α=70	*True Value*	-0.03	0.53	0	-0.03	0.53	0.06
		Mean	-0.03	0.37	0.04	-0.03	0.38	0.10
		Percent(%)	4.2	22.7	29.9	4.6	19.6	25.1

The number of sites matters to some extent. More sites tend to generate more bias. Although the bias of estimates is a little bit bigger in the situation of 5 sites, the probabilities to reject the true values are much bigger in the situation of 15 sites, especially for $\widehat{\beta_3}$, and by almost two or three times. If we compare the correlations between facilities and quality from Table 4-2 in the two situations, we can see that with the same α, this correlation is higher with fewer sites. When there are more sites with fixed number of recreationists, there is more variation in facilities. The aggregation effect is reduced, and individual factors have more influence over the supply of facilities. Their economic endogeneity is more likely to cause bias. So the results are consistent with the argument in Section 3.

Table 4-2. Correlations in Sensitivity Analyses

		TC-F	*Q-F*
5 Sites	*α=25*	-0.01	0.55
	α=70	-0.02	0.95
15 Sites	*α=25*	0	0.32
	α=70	0	0.67
Discrete F	*α=25*	0	0.50
	α=70	0	0.79
Random Parameter 1	*α=25*	0	0.26
	α=70	0	0.55
Random Parameter 2	*α=25*	-0.02	0.86
	α=70	0	0.86
Random Parameter 3	*α=25*	0.01	0.62
	α=70	0	0.87

Although the correlation between facilities and quality is lower in the situation with 15 sites than with 5 sites, it increases as α gets bigger. As before, this correlation needs to be very big to cause substantial bias in the estimates. In the situation with discrete facilities, the correlations also have a similar pattern. In the situations with different true parameters, the correlation still grows when α increases, but the magnitude of correlations is different, which may be attributed to the change of the relative importance of variables in the utility equation. A higher parameter means that variable becomes relatively more important. So when the true parameter of the quality variable is larger, the level of its correlation with facilities increases. The smaller β_2 is, the lower correlation that is needed to make the bias in the results substantial.

Since the level of correlations between facilities and site quality that cause bias to estimates differs with different number of sites and different parameters, a general rule for judging whether there will be statistical endogeneity by simply checking how the facility and quality variables are correlated is difficult to offer; however, it is still the case that the coefficient in the facility supply equation determines whether the economic endogeneity of facilities is a big problem.

Overall, the patterns observed in the above simulations appear robust for the types of sensitivity studies conducted here. Thus we continue using previous settings for the following simulations.

Section 5

SIMULATIONS WITHOUT FACILITY VARIABLE

Based on the results above, we notice one point: when the endogeneity causes bias to the estimated coefficient for facilities, the estimated coefficient for quality will also be affected. Thus we ask, will the endogeneity still influence that estimate when there is no facility variable in the recreation demand model? That is, can one simply drop the facilities variables? To test this, we do the simulations without the facility variable. That is, keeping other steps described in Section 2 to be the same, we regress people's choices only on travel cost and quality, even if there are newly built facilities available.

As we would expect, in Case I, no matter how large α is, the estimates are unaffected, as the probabilities to reject their true values are about 5%. There is no bias in the means and the standard errors. For both estimates, the variance and MSE are almost the same. When people don't care about facilities, it is correct to not include the facility variable.

In Case II, when people care about facilities but we do not put the facility variable in our regression, even though travel cost and quality are exogenous, we can see a big problem here. Again, $\widehat{\beta_1}$ is hardly affected. It remains unbiased with the percents around 5%. The means are equal to its true value in all situations, and the variance and MSE are also the same. $\widehat{\beta_2}$ now has an upward bias, which grows very fast as α increases. When α is 200, the mean is almost twice the true value. The variance and MSE become larger, too, and their differences are also getting bigger. The probability to reject its true value grows dramatically, from 9.8% to 99.3%. $\widehat{\beta_2}$ is biased in almost all the time when the economic endogeneity of facilities is very strong.

Table 5-1. Simulations without the Facility Variable

		Case I		*Case II*	
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$
α=10	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.52
	Var.	4.7e-06	6.23e-03	4.6e-06	8.2e-03
	MSE.	4.7e-06	6.23e-05	4.6e-06	8.8e-03
	*Percent(%)**	4.7	5.2	4.9	9.8
α=25	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.54
	Var.	4.1e-06	6.3e-03	4.2e-06	9.5e-03
	MSE.	4.1e-06	6.3e-03	4.2e-06	0.012
	Percent(%)	4.6	4.6	4.8	14.1
α=45	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.59
	Var.	4.4e-06	4.6e-03	4.4e-06	7.9e-03
	MSE.	4.4e-06	4.6e-03	4.4e-06	0.018
	Percent(%)	5.9	5.6	5.3	36.8
α=70	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.65
	Var.	4.5e-06	6.7e-03	4.4e-06	0.013
	MSE.	4.5e-06	6.7e-03	4.4-06	0.039
	Percent(%)	4.7	5.0	4.5	47.2
α=100	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.69
	Var.	4.8e-06	4.8e-03	4.9e-06	0.012
	MSE.	4.8e-06	4.8e-03	4.9e-06	0.054
	Percent(%)	7.0	6.3	6.6	76.0
α=200	*True Value*	-0.06	0.49	-0.06	0.49
	Mean	-0.06	0.49	-0.06	0.95
	Var.	4.6e-06	3.7e-03	4.8e-06	0.015
	MSE.	4.6e-06	3.7e-03	4.8e-06	0.23
	Percent(%)	5.0	5.2	5.8	99.3

The correlations between facilities and the two variables are shown in Table 5-2. Since the supply mechanism is the same as the previous simulations with endogeneity, the correlations should not be that different. Compared with those in Table 3-2, the results in Table 5-2 are similar, showing that the economic endogeneity is getting stronger as α increases.

Table 5-2. Correlations of Facilities with Travel Cost and Quality

	α=10	*α=25*	*α=45*	*α=70*	*α=100*	*α=200*
TC-F*	0	0	0	-0.01	0.01	-0.01
Q-F**	0.18	0.39	0.66	0.73	0.81	0.92

As long as facilities have impacts on people's choices, no matter whether the facility variable is included in the model or not, their economic endogeneity will influence the estimated coefficient of site quality. Actually, compared with previous results, the bias is a lot larger and in a different direction. Chances to reject the true value of $\widehat{\beta_2}$ are also much bigger when we don't put facilities in the regression in Case II. Statistical endogeneity comes from the fact that regressors are correlated with errors. The economic endogeneity of facilities makes facilities correlated with the error term in the utility equation. In Case II of basic simulations, facilities are in the regression, so the correlation between regressors and errors is just the correlation between facilities and errors, which is not very big because facilities are endogenous at the aggregate level. The bias in the estimates mainly comes from the induced multicollinearity of facilities with site quality. In Case II of the simulations above, facilities become part of errors. Since facilities are highly correlated with quality, now the correlation between regressors and errors is much larger which explains why the bias is more remarkable in this case.

Section 6

SIMULATIONS IN DIFFERENT LANDSCAPES

The landscape we have used does not involve any spatial clustering of individuals which implies maximal variation in the individual specific travel costs. Both people and recreation sites are spread out. Thus, the high variation in travel costs leads to very robust estimation of the travel cost parameter despite the level of facilities endogeneity. Also, 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 about 0.10 to be visited. 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. Therefore, the landscape could also have effects on the simulation results.

On the contrary, it would be common that recreationists cluster at some areas, like cities. Further, suppose sites are dispersed along a shoreline, such as beaches, rather than being randomly dispersed across the landscape. Now, by construction, there would be some sites that are more frequently visited than others. In fact, this situation is somewhat closer to reality. Based on empirical observations on the distribution of site visits, Lupi and Feather (1998) proposed an aggregation approach for recreation sites based on their popularity and potential for being altered by policy. 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 these cases, even if we focus on average past visitation, the popular sites would be more similar across choice sets.

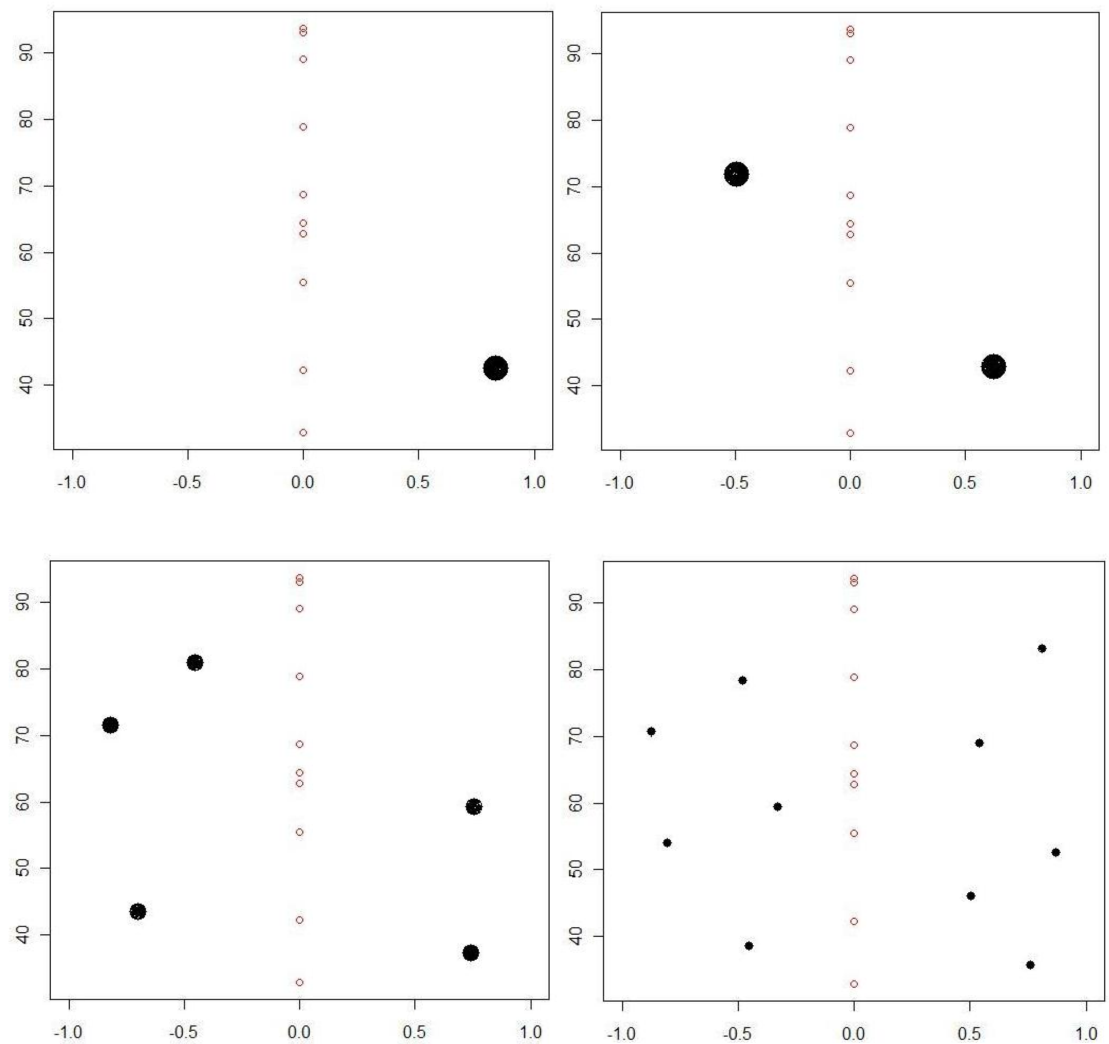

Figure 6-1. Four Landscapes with Clustering

Now we randomly draw 10 points on y-axis as recreation sites, and take random points inside one or more circles as people living in cities. The landscapes are shown in Figure 6-1. Simulations are run under the four landscapes, with other settings the same as before. The results are shown in Table 6-1.

Table 6-1. Simulation Results under Different Landscapes

			Case I			Case II		
			$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
One City	*True Value*		-0.06	0.49	0	-0.06	0.49	0.06
	α=25	*Mean*	-0.06	0.49	0	-0.06	0.49	0.06
		Var.	2.1e-05	7.8e-03	2.3e-04	3.9e-05	0.012	3.5e-04
		MSE.	2.1e-05	7.8e-03	2.3e-03	3.9e-05	0.012	3.5e-04
		%	5.2	5.8	5.2	4.6	6.0	4.9
	α=70	*Mean*	-0.06	0.49	0	-0.06	0.50	0.06
		Var.	4.4e-05	0.018	6.9e-05	1.7e-04	0.077	2.2e-04
		MSE.	4.4e-05	0.018	6.9e-05	1.7e-04	0.077	2.3e-04
		%	4.7	5.2	5.1	5.2	4.5	5.3
Two Cities	*True Value*		-0.06	0.49	0	-0.06	0.49	0.06
	α=25	*Mean*	-0.06	0.47	0.02	-0.06	0.47	0.08
		Var.	2.1e-04	7.1e-03	1.6e-03	2.2e-04	7.4e-03	1.6e-03
		MSE.	2.1e-04	7.4e-03	2.0e-03	2.2e-04	7.6e-03	2.0e-03
		%	5.1	4.8	8.0	5.3	5.1	8.0
	α=70	*Mean*	-0.06	0.44	0.01	-0.06	0.45	0.07
		Var.	5.3e-05	6.8e-03	3.5e-04	5.7e-05	7.1e-03	3.5e-04
		MSE.	6.9e-05	9.1e-03	5.5e-04	7.0e-05	8.9e-03	4.9e-04
		%	10.2	10.1	12.4	8.6	9.6	11.2
Five Cities	*True Value*		-0.06	0.49	0	-0.06	0.49	0.06
	α=25	*Mean*	-0.06	0.46	0.02	-0.06	0.46	0.08
		Var.	1.3e-05	6.6e-03	1.2e-03	1.3e-05	6.8e-03	1.2e-03
		MSE.	1.3e-05	7.3e-03	1.5e-03	1.3e-05	7.4e-03	1.5-03
		%	4.6	7.0	7.4	4.4	6.3	6.1
	α=70	*Mean*	-0.06	0.43	0.02	-0.06	0.43	0.07
		Var.	1.3e-05	7.5e-03	3.4e-04	1.4e-05	8.8e-03	3.9e-04
		MSE.	1.4e-05	0.011	6.1e-04	1.4e-05	0.012	6.0e-04
		%	4.7	10.9	14.8	4.6	9.2	10.4
Ten Cities	*True Value*		-0.06	0.49	0	-0.06	0.49	0.06
	α=25	*Mean*	-0.06	0.46	0.02	-0.06	0.46	0.08
		Var.	1.6e-05	7.1e-03	1.5e-03	1.6e-05	7.4e-03	1.5e-03
		MSE.	1.6e-05	7.9e-03	1.9e-03	1.6e-05	8.2e-03	1.9e-03
		%	4.8	6.8	9.6	5.3	6.0	8.8
	α=70	*Mean*	-0.06	0.34	0.04	-0.06	0.35	0.10
		Var.	1.1e-05	0.015	7.5e-04	1.1e-05	0.017	8.3e-04
		MSE.	1.1e-05	0.038	2.5e-03	1.1e-05	0.037	2.3e-03
		%	4.9	19.4	26.4	4.6	17.8	22.4

When we increase the coefficient in facility supply equation, we expect the bias to be more and more substantial, as in the previous landscape. However, this is not the case in all the four landscapes. In the landscapes with five and ten cities, we can see the chances to reject the true values are growing and the bias in the estimates is getting bigger as the coefficient increases. This makes sense because when the number of cities increases, the landscape becomes closer to the previous landscape where both people and recreation sites are randomly located. In the landscapes with one and two cities, especially when there is only one city, even if the coefficient is very large, the results are not very different from those with the coefficient equal to 25. The bias is very close to zero; the variance and MSE are almost the same; the probabilities to reject the estimates' true values are around 5% at 5% significance level. In other words, we don't really see a problem with economically endogenous facilities in the landscape with only one city.

Table 6-2. Correlations under Different Landscapes

	One City		*Two Cities*		*Five Cities*		*Ten Cities*	
	α=25	*α=70*	*α=25*	*α=70*	*α=25*	*α=70*	*α=25*	*α=70*
TC-F	-0.72	-0.68	-0.02	-0.24	-0.07	-0.16	-0.09	0.01
Q-F	0.42	0.25	0.39	0.56	0.54	0.84	0.54	0.85

From the correlations, if we look at the last three big columns, correlations between facilities and quality grow as α increases, and also more statistical endogeneity comes out. In the landscape with only one city, the correlation between facilities and travel cost is as big as -0.6 or -0.7. This is because all people face similar travel cost besides quality and facilities. Given travel cost is the most important in the decision-making process, nearly all people go to the same site. Past visitation highly relies on travel cost, as we mentioned in the beginning that travel cost is the most important in the utility equation, or to say, people's decision-making process. Even if a distant site has very good quality, most people won't go there. Thus, facilities are more likely to correlate with travel cost than with quality in this case. However, this big correlation does not cause bias to the estimates at all, even including the estimated coefficient for travel cost! Comparing it with the fact that a big correlation between facilities and quality leads to biased estimates, we would suspect that when people have almost the same choices, the effect of travel cost doesn't go away when we derive the average past visitation. Facilities have little influence on people's choices because they are built on sites which are already preferred by most people, so even a strong economic endogeneity won't cause big bias to the estimates.

To further prove this point, we create an extreme case: reducing the scale of travel cost to the same level as quality and facilities. Simulations before with only one city have the average value of travel cost about 150. The average values of endogenous facilities differ according to the coefficient in facility supply equation. They could be as low as 5 and as big as 100. In the following simulations, we set the average value of travel cost to be around 5. Then as the coefficient in facility supple equation increases, the scale of facilities become larger and facilities will dominate travel cost in the utility equation. The results are shown in Table 6-3 and Table 6-4.

For $\widehat{\beta_2}$ and $\widehat{\beta_3}$, when α is equal to 10, the results are very similar to previous ones. Given the scales of travel cost and facilities are more or less the same, when all people cluster in one city, we could still get some bias in the estimates for quality and facilities. The variance and MSE are more or less equal. When α goes to 25, 45 or even bigger, the bias increases, the probabilities to reject their true values increases and the differences between variance and MSE also become larger.

Table 6-3. Simulation Results under the Landscape with One City

		Case I			*Case II*		
		$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$
α=10	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.48	0	-0.06	0.48	0.07
	Var.	1.9e-04	5.7e-03	1.9e-03	1.9e-04	5.8e-03	2.0e-03
	MSE.	1.9e-04	5.8e-03	2.0e-03	1.9e-04	5.9e-03	2.1e-03
	Percent(%)	4.4	6.1	4.7	3.7	6.5	6.3
α=25	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.06	0.45	0.02	-0.06	0.46	0.08
	Var.	9.9e-04	8.1e-03	1.9e-03	1.0e-03	7.8e-03	1.9e-03
	MSE.	1.0e-03	9.3e-03	2.5e-03	1.0e-03	9.0e-03	2.5-03
	Percent(%)	4.0	7.1	9.1	3.4	6.5	8.6
α=45	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.05	0.41	0.04	-0.05	0.42	0.09
	Var.	2.5e-04	9.8e-03	1.3e-03	2.6e-04	0.010	1.3e-03
	MSE.	3.3e-04	0.016	2.6e-03	3.3e-04	0.016	2.5e-03
	Percent(%)	6.9	8.7	16.8	6.2	8.9	14.9
α=70	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.05	0.33	0.04	-0.05	0.34	0.10
	Var.	4.4e-04	0.024	8.2-04	5.4e-04	0.029	1.0e-03
	MSE.	6.3e-04	0.049	2.4e-03	6.9e-04	0.051	2.3e-03
	Percent(%)	5.5	12.1	23.7	4.7	12.4	18.0
α=100	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.04	0.31	0.03	-0.04	0.32	0.09
	Var.	2.8e-04	0.17	5.2e-04	3.3e-04	0.021	6.2e-04
	MSE.	7.4e-04	0.048	1.6e-03	7.6e-04	0.050	1.5e-03
	Percent(%)	19.4	25.3	28.8	15.7	21.0	19.5
α=200	*True Value*	-0.06	0.49	0	-0.06	0.49	0.06
	Mean	-0.02	0.15	0.03	-0.03	0.21	0.09
	Var.	1.9e-04	0.012	1.0e-04	4.3e-04	0.026	2.2-04
	MSE.	2.0e-03	0.125	1.3e-03	1.6e-03	0.103	9.3e-04
	Percent(%)	41.7	47.2	56.3	25.3	30.0	33.9

For $\widehat{\beta_1}$, when α is 10, there is no problem. When α goes to 25, although the mean is still the same as its true value, MSE starts to differ from the variance, implying that we might see a bias in the mean if we have a bigger α. When α is 45, the mean is greater than the true value by 0.01, a 17% upward bias. The percent of iterations we can reject its true value is a little larger than 5%, but still less than 10%. MSE now exceeds the variance by about 40%. As α keep increasing, the bias becomes bigger, and MSE is almost ten times of the variance in Case I with α equal 200. Nearly half of the time we can reject the true value of $\widehat{\beta_1}$.

Table 6-4. Correlations of Facilities with Travel Cost and Quality

	α=10	*α=25*	*α=45*	*α=70*	*α=100*	*α=200*
TC-F	-0.10	-0.03	-0.20	-0.83	-0.60	-0.61
Q-F	0.22	0.53	0.60	0.89	0.79	0.76

Therefore, under the landscape of only one city, even if the coefficient in the facility supply equation is very big and leads to a large correlation between facilities and travel cost, the results may not be very remarkable. When facilities become the most important preference variables in place of travel cost, their economic endogeneity causes problems, not only to the estimates of facilities and quality, but also to the estimate of travel cost.

Section 7

WELFARE EFFECTS

As stated in the introduction, whether the welfare estimates are biased or not has very important policy implications. Simulations make it possible to compute the true welfare measures. Kling (1988) compared the estimated welfare measures with the true welfare measures calculated from simulated data to examine the reliability of welfare estimates from recreation demand models. In our simulations, beyond the estimates for the parameters, we also calculate the true and estimated welfare measures under the basic landscape with random location and the landscape with only one city where the scale of travel cost is reduced. In the basic landscape, as the economic endogeneity of facilities gets stronger, there is a downward bias in $\widehat{\beta_2}$ and an upward bias in $\widehat{\beta_3}$; $\widehat{\beta_1}$ is unbiased. Simulations under other circumstances except the extreme case give similar results. In the extreme situation, when $\widehat{\beta_2}$ and $\widehat{\beta_3}$ are affected, there is also some upward bias in $\widehat{\beta_1}$. We want to see how the bias in estimated coefficients will affect the welfare estimates in the two cases.

Following the formulas stated in Section 2, two types of values are considered: site value and marginal value for environmental quality change. For each random sample, the individual estimates are averaged across individuals to obtain the welfare estimates for each site. Then we average these estimates across 1,000 iterations and use those averages to compare with the true values. In addition, we compute the parameter ratio of β_2/β_1, both true values and estimates.

Table 7-1. Site Value in Basic Simulations with Endogenous Facilities

#Site			*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
α=10	*Case I*	*WTP*	2.43	1.86	2.97	1.20	1.10	2.18	1.31	2.65	1.81	2.74
		$\widehat{WTP}$	2.43	1.84	2.98	1.20	1.11	2.18	1.31	2.64	1.81	2.75
	Case II	*WTP*	2.45	1.84	3.03	1.18	1.07	2.18	1.28	2.68	1.80	2.79
		$\widehat{WTP}$	2.45	1.83	3.03	1.17	1.08	2.18	1.29	2.67	1.79	2.79
α=25	*Case I*	*WTP*	1.53	1.42	1.52	2.97	2.71	2.87	2.39	1.53	1.32	1.93
		$\widehat{WTP}$	1.52	1.42	1.51	2.97	2.72	2.87	2.39	1.52	1.32	1.95
	Case II	*WTP*	1.47	1.35	1.46	3.12	2.80	3.01	2.42	1.47	1.26	1.91
		$\widehat{WTP}$	1.46	1.36	1.45	3.11	2.81	2.99	2.42	1.46	1.26	1.93
α=45	*Case I*	*WTP*	2.21	3.19	1.61	1.54	1.23	1.78	2.64	1.74	2.50	1.77
		$\widehat{WTP}$	2.21	3.23	1.62	1.52	1.22	1.77	2.69	1.74	2.45	1.77
	Case II	*WTP*	2.25	3.55	1.53	1.44	1.11	1.72	2.80	1.68	2.60	1.71
		$\widehat{WTP}$	2.24	3.57	1.53	1.42	1.11	1.72	2.83	1.68	2.54	1.70
α=70	*Case I*	*WTP*	1.60	2.25	2.26	1.87	1.67	1.37	2.78	2.55	1.95	1.84
		$\widehat{WTP}$	1.62	2.30	2.22	1.85	1.67	1.57	2.83	2.53	1.97	1.81
	Case II	*WTP*	1.48	2.33	2.33	1.82	1.57	1.22	3.10	2.75	1.92	1.78
		$\widehat{WTP}$	1.51	2.37	2.29	1.79	1.57	1.22	3.14	2.72	1.93	1.75
α=100	*Case I*	*WTP*	1.79	1.53	1.93	1.26	2.08	2.43	2.96	2.71	1.67	1.84
		$\widehat{WTP}$	1.77	1.56	1.96	1.27	2.02	2.43	3.02	2.69	1.63	1.85
	Case II	*WTP*	1.67	1.32	1.86	1.02	2.06	2.61	3.62	3.10	1.50	1.73
		$\widehat{WTP}$	1.65	1.34	1.88	1.03	2.00	2.61	3.65	3.06	1.47	1.73
α=200	*Case I*	*WTP*	1.39	2.68	1.97	1.80	2.50	1.42	3.04	1.47	1.50	2.44
		$\widehat{WTP}$	1.44	2.76	1.87	1.70	2.49	1.42	3.11	1.44	1.55	2.43
	Case II	*WTP*	0.95	3.44	1.82	1.54	2.89	1.00	4.40	1.06	1.09	2.83
		$\widehat{WTP}$	0.99	3.48	1.73	1.45	2.86	0.99	4.43	1.04	1.13	2.80

Based on the results in Table 7-1, estimated site values are not very different from the true values. Only one or two deviate from the truth by more than 10%. Most are greater or smaller than their true values by around 2%, even in the case with quite remarkable bias in $\widehat{\beta_2}$ and $\widehat{\beta_3}$. This is not surprising though. There is a downward bias in $\widehat{\beta_2}$, but an upward bias in $\widehat{\beta_3}$. Maximum likelihood seeks parameters that are most likely to generate the observations, so although the estimates are different, the predicted

Table 7-2. Site Value in the Landscape with One City

	#Site		1	2	3	4	5	6	7	8	9	10
α=10	Case I	WTP	2.01	1.15	1.58	1.91	1.55	1.19	2.77	1.69	1.33	2.44
		$\widehat{WTP}$	2.15	1.26	1.69	2.01	1.66	1.29	2.95	1.83	1.43	2.60
	Case II	WTP	2.02	1.12	1.56	1.91	1.53	1.17	2.85	1.68	1.31	2.49
		$\widehat{WTP}$	2.17	1.22	1.67	2.02	1.64	1.26	3.05	1.81	1.40	2.65
α=25	Case I	WTP	1.69	1.43	1.43	1.40	1.42	2.27	1.76	2.96	1.24	2.04
		$\widehat{WTP}$	3.23	2.58	2.77	2.60	2.62	4.05	3.29	5.57	2.24	3.87
	Case II	WTP	1.67	1.37	1.38	1.34	1.37	2.34	1.74	3.25	1.17	2.08
		$\widehat{WTP}$	1.57	1.30	1.27	1.19	1.17	2.19	1.56	3.00	1.11	1.93
α=45	Case I	WTP	2.16	1.34	1.68	1.68	1.69	1.94	2.52	2.40	1.17	1.05
		$\widehat{WTP}$	3.01	1.90	2.39	2.38	2.33	2.68	3.55	3.36	1.69	1.51
	Case II	WTP	2.26	1.24	1.63	1.63	1.64	1.95	2.77	2.61	1.05	0.93
		$\widehat{WTP}$	2.52	1.34	1.83	1.92	1.75	2.19	3.15	2.87	1.23	1.05
α=70	Case I	WTP	1.03	2.89	1.66	1.70	1.14	3.08	0.88	2.09	1.61	1.62
		$\widehat{WTP}$	1.46	3.99	2.24	2.22	1.62	4.30	1.29	2.79	2.12	2.15
	Case II	WTP	0.81	3.52	1.51	1.57	0.92	3.92	0.67	2.11	1.45	1.46
		$\widehat{WTP}$	1.73	7.28	3.11	3.23	2.71	8.27	1.49	4.31	2.54	2.68
α=100	Case I	WTP	0.86	2.02	2.04	1.36	1.49	1.05	2.14	1.43	3.36	1.93
		$\widehat{WTP}$	1.86	4.01	4.16	2.86	3.15	2.25	4.40	2.84	7.13	3.84
	Case II	WTP	0.57	1.99	2.03	1.08	1.23	0.75	2.18	1.16	5.28	1.83
		$\widehat{WTP}$	1.01	3.19	3.44	1.61	2.01	1.28	3.06	1.91	8.42	2.73
α=200	Case I	WTP	2.58	1.19	1.29	1.72	2.45	2.57	1.82	1.08	1.67	1.27
		$\widehat{WTP}$	47.4	23.8	22.3	32.4	49.3	47.3	34.7	21.5	30.0	22[1]
	Case II	WTP	3.85	0.69	0.80	1.42	3.33	3.78	1.61	0.59	1.33	0.78
		$\widehat{WTP}$	-22	-3.5	-5.1	-12	-8.4	-22	-6.8	-3.1	-3.5	-2.8[2]

[1] In Case I when α=200, $\widehat{\beta_1}$ is severely biased. In total, 86 in 1,000 iterations produce positive $\widehat{\beta_1}$, the maximum of which is 0.04. Particularly, there are 30 $\widehat{\beta_1}$s distributed within (-0.001, 0.001), among which 18 is negative and 12 is positive. The estimated WTPs correspondingly are extremely large. The negative $\widehat{\beta_1}$s dominate the positive ones, so the average WTPs are positive. Take the estimated WTP for site 1 as an example, among the 1,000 WTPs, the minimum is -3141, the maximum is 17,440. The median is 7.93.

[2] In Case II, 99 in 1,000 iterations give positive $\widehat{\beta_1}$, the maximum of which is 0.05. There are 17 $\widehat{\beta_1}$s distributed within (-0.001, 0.001), among which 12 is positive. They generate very large and negative WTPs, As they dominate the negative $\widehat{\beta_1}$s, the average WTPs are negative in this situation. For the estimated WTP of site 1, the maximum is 670.9 and the minimum is -23,650! The median is 7.54 by the way, so although it is rare to have more positive $\widehat{\beta_1}$s within (-0.001, 0.001) in consideration of its true value, the overall distribution of $\widehat{\beta_1}$ does not go wrong. It is those extreme values that flip the sign of the averages. We re-run simulations under this situation for another two times. Both have positive estimated WTPs in Case II. In one simulation some of the estimated WTPs in Case I and Case II are still as large as 20; in the other simulation, those estimates go back to the true values' magnitude and are around 6 or 7, close to the medians. Anyway, those negative estimated WTPs are unusual, but that might happen.

probability of visiting each site is similar. Plus, there is no bias in $\widehat{\beta_1}$, which appears in the denominator of the site valuation equation. Therefore the estimated site values are pretty close to the truth.

Table 7-3. Marginal Value of Quality Change in Basic Simulations

			*R***	*1****	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
α=10	*Case I*	*T**	-8.2	0.96	0.77	1.13	0.51	0.48	0.89	0.56	1.03	0.76	1.07
		*E**	-8.1	0.96	0.76	1.13	0.50	0.48	0.88	0.56	1.03	0.75	1.07
	Case II	*T*	-8.2	0.97	0.76	1.15	0.50	0.47	0.89	0.55	1.04	0.75	1.09
		E	-8.1	0.97	0.76	1.15	0.49	0.47	0.88	0.55	1.04	0.75	1.09
α=25	*Case I*	*T*	-8.2	0.64	0.60	0.64	1.14	1.06	1.12	0.95	0.65	0.58	0.79
		E	-7.8	0.61	0.58	0.61	1.10	1.02	1.07	0.91	0.62	0.55	0.76
	Case II	*T*	-8.2	0.62	0.58	0.62	1.19	1.09	1.16	0.96	0.63	0.55	0.78
		E	-7.8	0.59	0.55	0.59	1.14	1.05	1.11	0.92	0.60	0.52	0.75
α=45	*Case I*	*T*	-8.2	0.90	1.21	0.67	0.64	0.53	0.74	1.03	0.72	0.98	0.73
		E	-7.1	0.78	1.06	0.58	0.55	0.46	0.64	0.90	0.62	0.84	0.64
	Case II	*T*	-8.2	0.91	1.32	0.64	0.60	0.49	0.72	1.08	0.70	1.01	0.71
		E	-7.1	0.79	1.16	0.56	0.52	0.42	0.63	0.95	0.61	0.88	0.62
α=70	*Case I*	*T*	-8.2	0.67	0.90	0.90	0.77	0.70	0.59	1.07	1.00	0.79	0.77
		E	-6.3	0.52	0.70	0.68	0.59	0.53	0.45	0.84	0.77	0.61	0.58
	Case II	*T*	-8.2	0.63	0.92	0.92	0.75	0.66	0.53	1.17	1.06	0.78	0.74
		E	-6.3	0.49	0.73	0.71	0.58	0.51	0.41	0.93	0.82	0.61	0.57
α=100	*Case I*	*T*	-8.2	0.75	0.63	0.79	0.54	0.84	0.96	1.15	1.06	0.70	0.76
		E	-4.8	0.43	0.38	0.47	0.32	0.48	0.57	0.69	0.62	0.40	0.45
	Case II	*T*	-8.2	0.70	0.55	0.76	0.44	0.82	1.01	1.34	1.18	0.63	0.71
		E	-5.1	0.43	0.35	0.48	0.27	0.50	0.63	0.85	0.73	0.38	0.44
α=200	*Case I*	*T*	-8.2	0.59	1.05	0.81	0.74	0.98	0.60	1.16	0.62	0.64	0.97
		E	-3.8	0.28	0.50	0.36	0.32	0.45	0.28	0.55	0.28	0.30	0.45
	Case II	*T*	-8.2	0.42	1.28	0.74	0.63	1.10	0.43	1.54	0.46	0.48	1.09
		E	-4.3	0.22	0.68	0.37	0.32	0.58	0.22	0.83	0.23	0.25	0.57

*T: True value; E: Estimated value; **R: Ratio of β_2/β_1; ***1: Site number

On the contrary, in landscape with only one city where the scale of travel cost is relatively small, as α goes up, there is a huge bias in the estimates, which could even be as large as the estimates themselves. We also see problems in β_1, so the estimated site values are no longer reliable.

Table 7-4. Marginal Value of Quality Change in the Landscape with One City

			R	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
α=10	*Case I*	*T*	-8.2	0.93	0.55	0.74	0.88	0.72	0.56	1.25	0.79	0.63	1.11
		E	-8.5	0.97	0.57	0.77	0.90	0.75	0.59	1.30	0.83	0.65	1.15
	Case II	*T*	-8.2	0.93	0.53	0.73	0.88	0.72	0.55	1.29	0.78	0.62	1.13
		E	-8.5	0.97	0.56	0.75	0.90	0.74	0.58	1.34	0.82	0.64	1.18
α=25	*Case I*	*T*	-8.2	0.79	0.67	0.67	0.66	0.66	1.04	0.82	1.33	0.58	0.94
		E	-14	1.35	1.09	1.16	1.10	1.10	1.68	1.38	2.26	0.95	1.62
	Case II	*T*	-8.2	0.78	0.64	0.65	0.64	0.64	1.06	0.81	1.45	0.55	0.95
		E	-7.5	0.73	0.60	0.60	0.56	0.55	1.00	0.73	1.35	0.52	0.89
α=45	*Case I*	*T*	-8.2	0.99	0.63	0.78	0.78	0.79	0.90	1.14	1.10	0.56	0.50
		E	-8.8	1.07	0.68	0.85	0.85	0.83	0.96	1.25	1.18	0.61	0.54
	Case II	*T*	-8.2	1.03	0.59	0.76	0.76	0.76	0.90	1.24	1.18	0.50	0.44
		E	-8.1	1.02	0.57	0.75	0.78	0.73	0.90	1.25	1.16	0.51	0.44
α=70	*Case I*	*T*	-8.2	0.48	1.30	0.77	0.80	0.54	1.38	0.42	0.96	0.75	0.76
		E	-8.1	0.49	1.30	0.75	0.77	0.55	1.41	0.44	0.92	0.74	0.72
	Case II	*T*	-8.2	0.38	1.55	0.71	0.74	0.44	1.71	0.32	0.97	0.68	0.68
		E	-11	0.53	2.02	0.90	0.92	0.76	2.30	0.44	1.21	0.75	0.80
α=100	*Case I*	*T*	-8.2	0.41	0.93	0.95	0.64	0.70	0.50	0.98	0.67	1.50	0.89
		E	-9.3	0.49	1.04	1.07	0.74	0.81	0.58	1.10	0.75	1.76	0.98
	Case II	*T*	-8.2	0.28	0.91	0.93	0.51	0.59	0.36	0.99	0.55	2.20	0.85
		E	-7.9	0.29	0.88	0.93	0.48	0.58	0.36	0.90	0.54	2.18	0.78
α=200	*Case I*	*T*	-8.2	1.17	0.56	0.61	0.80	1.12	1.16	0.85	0.52	0.78	0.60
		E	-33	4.68	2.42	2.27	3.27	4.83	4.67	3.45	2.20	3.03	2.26
	Case II	*T*	-8.2	1.66	0.33	0.38	0.67	1.46	1.64	0.75	0.28	0.62	0.37
		E	-13	2.66	0.54	0.66	1.31	2.02	2.73	1.11	0.45	0.79	0.49

Table 7-3 and Table 7-4 present the marginal values for quality changes. Both results show that when the bias in the ratio of parameters grows, the difference between estimated values and true values are also getting bigger. In the basic landscape, if α is larger, the ratios of estimated quality parameters are smaller in absolute value, which is consistent with the simulation results. The estimated marginal value of quality change is also diminishing. Thus, although the economic endogeneity of facilities does not have much influence on the site value, it does affect the value of environmental quality.

In landscape with only one city where the scale of travel cost is reduced, the trend of the bias is not as obvious as that in the basic landscape. But if we compare the results with α equal 10 and 200, we can see a huge difference. Although both $\widehat{\beta_1}$ and $\widehat{\beta_2}$ have a downward bias, the ratio of their estimates still deviates from the truth. When all estimated coefficients in the utility equation are biased, the welfare estimates for quality change are also biased.

Section 8

CONCLUSIONS

Researchers who conduct travel cost studies of recreation demand are often confronted with site characteristics that are not exogenously determined. In these situations researchers might rightly be concerned about the potential implications of endogenous site characteristics. In this paper, we use Monte Carlo simulations to investigate whether the economic endogeneity of facilities causes bias to the coefficient estimates and welfare estimates. The answer to the research question is: the economic endogeneity of site facilities leads to statistical endogeneity, which means the facility variable is correlated with individual errors, but the statistical endogeneity is not strong enough in most cases to cause a big problem in the estimation since facilities are endogenous at the aggregate level. However, this economic endogeneity still needs attention, because it also can induce multicollinearity of facilities with other site characteristics, which may cause bias to the estimates. Based on the answer, a few conclusions can be drawn from our simulations:

(1) In the situation when there is a lot of variation in travel cost, which is identical to the basic landscape we used in the simulations, the strength of facilities' economic endogeneity matters, because strong economic endogeneity results in high multicollinearity. If the endogeneity is very strong, we would expect a bias in the estimated coefficients of facilities and other site characteristic variables; otherwise, there is no need to worry about it. We can calculate the correlation between facilities and other site characteristic variables for diagnosis. The

correlation should be very outstanding to cause a big problem. The threshold of the correlation to lead to bias is affected by the underlying factors of the study, like the number of sites, the sample characteristics, etc. Sometimes it has to be as high as 0.7.

(2) No matter how strong the economic endogeneity is, there is almost always no bias in the estimated coefficient for travel cost, and we don't see a huge bias in the welfare measure of site value. As for the value of marginal quality change on sites, the bias is growing as the endogeneity gets stronger. Thus the site value is less sensitive and more reliable than the value of marginal quality change.

(3) When there is economic endogeneity of facilities, dropping the facility variable does not provide a clear solution: it will make matters worse when people do care about site facilities, but including the facilities can lead to spurious parameter significance when facilities are included yet people do not care about them.

(4) In the situation where people cluster, especially where all people live in one area, which means all recreationists have similar travel costs, we don't quite get bias in all estimates even if the endogeneity is very strong, except in extreme cases. We have to go to some length to construct simulations and landscapes which induce bias in the travel cost parameters. As the number of clusters increases, the results become closer to the situation described in the first conclusion.

The results above apply for not only facilities, but also other site characteristics that are provided by site managers or a third party. Examples of site quality characteristics that depend on some endogenous level of investment in monitoring or measurement and have been studied in recreational demand settings include water quality advisories for a beach, fish consumption advisories, etc. For such variables, we would expect managers to allocate their measurement and monitoring budget so that more popular sites are more likely to be measured/monitored.

Further, in this study, we assume quality to be the same for all people; in other words, we use objective quality measures. The estimated coefficient of quality is also influenced by the multicollinearity that can be induced through the economic endogeneity of the supply of facilities. Some studies valuing recreation demand adopt subjective quality measures rather than objective ones, especially in the single site model. If perceived quality data is used instead of objective data, quality variable will vary across both people *and* sites, as does travel cost. Thus, its correlation with facilities will greatly decrease, and we would not expect to see bias in the estimated coefficients.

In a word, in empirical studies, we need to be careful with the data and with site characteristics suspected to be economically endogenous, e.g., facilities. Even though they are site-specific, their economic endogeneity may still have substantial influence on the estimation results. As possible diagnostics for economic endogeneity problems, one can examine the distribution of recreation sites and recreationists, and compute the correlations of that particular site characteristic with other site characteristic variables. Then, based on the expected impact of the endogeneity and depending on the estimates the study is interested in, one can decide whether to take the possible economic endogeneity into account.

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

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(3): 370-381.

Dhar, T., J. P. Chavas, et al. (2003). "An empirical assessment of endogeneity issues in demand analysis for differentiated products." American Journal of Agricultural Economics 85(3): 605-617.

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.

Kling, C. L. (1988). "Comparing welfare estimates of environmental quality changes from recreation demand models." Journal of Environmental Economics and Management 15(3): 331-340.

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.

Villas-Boas, J. M. and R. S. Winer (1999). "Endogeneity in Brand Choice Models." Management Science 45(10): 1324-1338.

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): 1061-1076.

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

Yang, S., Y. Chen, et al. (2003). “Bayesian analysis of simultaneous demand and supply.” Quantitative Marketing and Economics 1(3): 251-275.

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.