

# EFFECTS OF SAMPLE SELECTION ON ESTIMATES OF ECONOMIC IMPACTS OF OUTDOOR RECREATION'

*Donald B.K. English \**

**Abstract**—Estimates of the economic impacts of recreation often come from spending data provided by a self-selected subset of a random sample of site visitors. The subset is frequently less than half the onsite sample. Biased vectors of per trip spending and impact estimates can result if self-selection is related to spending patterns, and proper corrective procedures are not employed. This paper shows a method for accounting for both sample selection and the censored nature of reported expenditures, via a Tobit model with sample selection. Results from a sample of visitors to Cumberland Island National Seashore indicate a naive (uncorrected) approach overestimates per trip visitor spending by 15 percent and economic impacts to industrial output by 10 percent.

## I. INTRODUCTION

Estimates of the economic impacts associated with recreation visitation are useful policy tools for many federal and state agencies in making land management decisions (Jackson, et al., 1992; Johnson and Moore, 1993; USDA Forest Service, 1995). Many times these agencies are as interested in the distributive effects of their policy decisions as they are in the welfare effects. Recreation has been touted as an environmentally friendly method of achieving rural development goals (Bergstrom, et al., 1990; Dawson, et al., 1993). As a result, recreation impacts can be of particular interest for Federal agencies such as the USDA-Forest Service that include economic development or diversity as an explicit consideration in planning processes and that manage significant amounts of recreation resources.

Approaches used in assessing economic impacts of recreation visitation have been reasonably well defined and documented (Alward and Lofting, 1985; Alward, et al., 1985; Stevens and Rose, 1985; Leitch and Leistriz, 1985; Johnson and Moore, 1993; English and Bergstrom, 1994). Recreation is generally treated as an export industry, so economic growth is due to spending by nonresident visitors. Impact estimates for visitation at a recreation site are generally derived by first identifying the target population(s) of users, and obtaining an estimate of

average spending per visitor per trip (or per visitor day) within the target economic region for a vector of trip-related items. Next, the average expenditure vector is treated as a final demand vector for an economic (often Input-Output) model to estimate per person impacts. Finally, per person per trip impacts are multiplied by an estimate of nonlocal site visitation for each relevant user type (**Propst**, et al., 1992; English and Bergstrom, 1994).

Visitor spending data are usually obtained from mailing a survey to a random sample of visitors who are first contacted while at the site. The mailing is timed so that the survey arrives soon after the visitor returns home. Expenditure surveys generally ask visitors to report the amount of money spent by their party in a given geographic area for food, lodging, gasoline, and other types of expenses associated with the recreation trip on which they were initially contacted. The number of **people** whose expenses are represented is also obtained.

Some studies recognize that visitors engaging in different activities have different spending patterns, so analyses are sometimes separated according to primary activity, and average expenditure vectors obtained for each user type (**Propst** et al., 1992; Schorr, 1995). Length of stay at the site may also be important, and this can be another dimension for developing separate spending profiles (Jackson, 1992; Propst, 1992). For people on multiple site trips, adjustments that prorate expenditures to the target site as a function of time spent at other sites or number of other sites visited are common (Cordell, et al., 1992).

Most studies in the journal literature do not report their actual calculation method, but simply state that they develop an estimate of trip spending per person (**Dawson**, et al., 1993; Cordell, et al., 1992). Usually, total party expenditures and party size are separate questions. Thus, the correct method to estimate the population ratio of spending per person is to divide average spending per party by the average party size, rather than take the average of spending per person **from** each sampled party (**Cochran**, 1977).

Unfortunately, response rates to the mailed expenditure surveys have varied from around 30 percent (Bergstrom, et al., 1990; Cordell, et al., 1990) to over 70 percent (**Dawson**, et al., 1993; Johnson and Moore, 1993). In some cases, the response rate is not known (Schorr, et al., 1995), although response rates between 40 and 60 percent seem to be the norm. As a result, for many studies the sample of visitors whose expenditures are used to develop spending profiles is a subset of the **onsite** random sample. More importantly, it is a subset for which the visitors themselves self-select for inclusion.

Such self-selection presents a possible source of bias in the visitor spending profiles. Self-selection in samples has been identified as an important issue in a number of applications in both sociology and economics (Maddala, 1983;

Winship and Mare, 1992). With recreation in particular, sample selection issues have been mostly concerned with the effects of exclusion of nonparticipants and/or endogenous stratification on estimates of consumer surplus (Shaw, 1988; Smith, 1988; Willis and Garrod, 1991; Dobbs, 1993). Heckman (1979) was among the first to note that sample selection can lead to biased inferences about social processes. In the present context, if the decision to respond to mailed expenditure surveys is not independent of expenditure levels or patterns, then standard practices can lead to biased mean spending profiles.<sup>2</sup>

Many studies 'correct' their expenditure samples by weighting to reflect the onsite population according to demographic variables such as race and gender. However, if the weighting variables are not those that are related to both amount spent and probability of returning the expenditure survey, then very little may be gained. Complications in developing corrective procedures may arise due to the nature of reported spending in a regional economy. Any individual item (for example, meals at restaurants) is often not purchased by a substantial portion of respondents. As a result, observed values for in region spending are likely to be censored at zero.

This paper examines the combined effects of sample selection and censoring on estimates of spending per person per trip and on subsequent economic impact estimates. Empirical data are used from visitors to Cumberland Island National Seashore (CINS) on Georgia's coast. In the next section, the models and data are described. The following section presents results and compares impact estimates from the unweighted expenditure sample (naive) means to those obtained when accounting for sample selection and/ or censoring. The final section discusses the implications of the results and makes recommendations for future research.

## II. MODEL

Consider the case of a typical survey process designed to obtain recreation visitors' expenditures to a particular site. The process begins with an onsite random sample of  $N$  visitors or visitor groups. All of those contacted provide information on some vector  $V$  of data items on themselves and/or their group, such as demographics, use patterns, or trip descriptors. In addition, all are mailed a survey which contains questions about their party's expenditures ( $Y$ ) for  $j$  items made near the recreation site.

Some number of respondents,  $M$  ( $M < N$ ), return the expenditure survey. Those who choose to return the survey do so because they feel the value of the information they provide is above some threshold level. For those who fail to return the survey, the information's value is below the threshold. Let  $Z^*$  be the

perceived value for returning the survey. The unobservable process that determines  $Z^*$  is presumed to be a **function** of information collected during the **onsite** contact. However, all that can be observed is whether the person returned the survey or not. Let  $Z$  be an indicator variable for the value of  $Z^*$  that equals 1 if an expenditure survey is returned and zero otherwise. Then, if the error term is assumed normal, the observable portion of the decision process can be modeled via a **probit** equation.

The amount of money spent near the recreation site is assumed to be explained by another vector of variables,  $X$ . Information on these variables can be obtained through either the **onsite** contact or mailed survey. This vector may be **no** more than a series of dummy variables describing primary recreation activities, or it may include variables thought to be related to expenditure levels, such as group size, travel distance, or length of stay. We can assume that the utility maximizing level of expenditures,  $Y^*$ , is a linear function of the variables in  $X$ , plus a normally distributed error term. However, spending less than zero dollars generally cannot happen. Thus, the observed level of expenditures,  $Y$ , for any item are censored at zero, and can be modeled via a **Tobit** equation. Of course, expenditures and perhaps some portion of  $X$  are observed only for those persons who return that **survey** (i.e., when  $Z=1$ ).

Assume that expenditures on any two items are independent, so that spending for each item can be modeled without regard to any of the other purchases. Then by combining the selection mechanism for returning the expenditure survey with the determination of the amount spent, the complete model for individual  $i$  and expenditure item  $j$  is:

$$\begin{aligned} Y_{ij}^* &= X_{ij}\beta_j + e_{ij} \\ Z_i^* &= V_i\alpha + \mu_i \end{aligned}$$

where  $e_{ij}$  and  $\mu_i$  are jointly normally distributed, with means of zero, variances  $(\sigma_j^2, 1)$ , and covariance  $\sigma_{j\mu}$  and where  $Y_{ij}$  and  $Z$  are observed rather than their starred counterparts. The relationships between the unobserved and observed variables are:

$$\begin{aligned} Y_{ij} &= Y_{ij}^* \text{ if } Y_{ij}^* > 0 \\ &= 0 \text{ otherwise;} \\ Z_i &= 1 \text{ if } Z_i^* > 0 \\ &= 0 \text{ otherwise;} \end{aligned}$$

The combination of the two equations yields a zero-censored **Tobit** model that includes sample selection. For such a model, the log-likelihood function for expenditure item  $j$  is (Nawata, 1994; Greene 1995 p.624):

$$\log L = \sum_{Z=0} \ln \Phi(-V\alpha) + \sum_{Z=1, Y_j=0} \ln \Phi_2\left(\frac{-X\beta_j}{\sigma_j}, V\alpha, \rho_j\right) + \sum_{Z=1, Y_j>0} \ln \phi(Y_j - X\beta_j) + \ln \Phi\left[\left(V\alpha + \frac{\rho_j(Y_j - X\beta_j)}{\sigma_j}\right) / (1 - \rho_j^2)^{1/2}\right] \quad (3)$$

- where  $\rho_j$  = correlation between  $\sigma_j$  and  $\mu$ ,
- $\Phi_2$  = bivariate standard normal cdf,
- $\Phi$  = standard normal cdf,
- $\phi$  = standard normal pdf.

In essence, the three terms in equation 3 indicate that to estimate the model, there are three possible conditions that must be evaluated simultaneously. First, there is the probability that the individual did not provide expenditure information ( $Z=0$ ). Second is the combined probability that the person did provide expenditure data, but the observed value was censored ( $Z=1, Y_j=0$ ). Finally, there is the probability that **nonzero** expenditure information was provided ( $Z=1, Y_j>0$ ). The equation can be estimated via maximum likelihood. If the estimated parameter value for  $\rho_j$  is not significantly different from zero, then sample selection is not likely to be a serious problem, and a simpler model can be used.

The conditional mean for  $Y_j$ , given that expenditures are reported, can be found in **Maddala** (1983, p. 368) or **Greene** (1995 p.603). That conditional mean is:

$$E[Y_j | Z=1] = \Phi_2\left(\frac{X\beta_j}{\sigma_j}, V\alpha, \rho_j\right)X\beta_j + \sigma_j \left[ \phi\left(\frac{-X\beta_j}{\sigma_j}\right)\Phi\left(\frac{V\alpha}{(1-\rho_j^2)^{1/2}}\right) + \rho_j \phi(-V\alpha)\Phi\left(\frac{\frac{X\beta_j}{\sigma_j} - \rho_j V\alpha}{(1-\rho_j^2)^{1/2}}\right) \right] \quad (4)$$

From equation 4, it is evident that sample selection bias has a potentially important role in determining the mean expenditure per group. The correlation

coefficient between the selection and regression equation affects all of the terms in the conditional mean.

In the typical survey process, expenditure information for the group and the size of the group for whom expenditures are reported (**PAYFOR**) are separate variables, both of which can vary across observations. However, the value of interest from the sample is expenditures per person, the ratio of the two variables. **Cochran** (1977, p.30) shows that the correct sample estimate of a ratio of two random variables is the ratio of the respective sample means. Thus, the estimate of per person per trip expenditures for item  $j$  ( $YP_j$ ) is:

$$YP_j = \frac{E[Y_j | Z=1]}{E[\text{PAYFOR}]} \quad (5)$$

where  $E[\text{PAYFOR}]$  is the average recreation group size, for the subsample who responded to the expenditure survey.

### III. DATA

Data were collected *using* an **onsite** (intercept) survey that contained questions about trip length, primary activity, annual use patterns, and group demographics. Expenditure surveys were mailed to all those contacted **onsite**, following **Dillman's** (1978) procedures. The expenditure surveys asked about trip related spending for three different parts of the trip: at home, while traveling, or while at the site. In addition, other questions gathered information on the number of people paid for, and the proportion of traveling spending made within 50 miles of the recreation site. There were 980 **onsite** (intercept) interviews conducted during the fall of 1993 through spring of 1994. Of these, 440 completed **mailback** expenditure surveys were returned (44.9 percent response).

Expenditure data on fourteen expenditure items were collected (Table 1). Spending near **CINS** included all money spent while at the site, and the reported proportion of travel spending made within 50 miles of CINS. For half of the expenditure items, fewer than 10 percent of respondents reported spending any money within 50 miles of CINS. These infrequently purchased items were not included in any further analysis. Five items, PVTLODGE, FOOD, RSTRNT, GAS, and BOAT, were purchased by over 40 percent of the respondents.

The intent of this effort was to examine estimates of spending per person per trip by nonlocals. Consequently, those people reporting one-way travel of less than 50 miles were deleted. A few other cases were deleted because data on travel distance, primary activity, or number of people for whom expenses were

TABLE 1  
Expenditure Items and Percentage of Respondents  
Making some Purchase within 50 miles of CMS

Expenditure Item (name)	Percent Purchasing
Food&inks bought at restaurants ( <b>RSTRNT</b> )	56.1
Gasoline and oil for automobiles (GAS)	50.0
Boat rental or ferry fares (BOAT)	47.1
Privately owned lodging ( <b>PVTLODGE</b> )	46.5
Food/drinks bought at stores (FOOD)	42.0
souvenirs ( <b>SOUV</b> )	14.9
Publicly owned lodging (PULODGE)	13.8
Admissions to tourist attractions ( <b>ADMISS</b> )	9.8
Other transportation costs ( <b>OTRAN</b> )	7.8
Film or film development (FILM)	7.7
Clothing (CLOTH)	4.3
Fishing guides, bait or tackle (FISH)	2.2
Personal or health services (SERVE)	1.6
Equipment (EQUIP)	1.6

paid were missing. The final sample included 812 **onsite** observations, of which 370 had provided expenditure data.

#### IV. EMPIRICAL MODELING

*Expenditure survey response.* Modeling proceeded in two stages. In the first stage, response to the expenditure survey was assumed to be a function of demographic variables, including age, race, gender, household structure, and education, as well as trip-related variables including distance traveled and length of stay at CINS. Income information was missing for nearly one-third of the **onsite** contacts. Because a conceptual link between income level and returning the expenditure survey was not clearly evident, and in order to keep the analysis manageable, income was not included as a **regressor**.<sup>3</sup>

Consistent with the model given above, a **probit** specification was used to examine the relationship. The purpose of this stage of the modeling was to determine which variables to include in the vector V for use in the maximum

likelihood estimation. Variables were retained if **significance** levels were less than 0.10. Although estimation for both stages was performed using **LIMDEP** 7.0 (Greene, 1995), any of a number of commercially available statistical software packages that allow maximization (or minimization) of user-defined functions could have been employed.

Six variables, plus a constant, were retained for inclusion in the second stage of modeling (Table 2). These variables included dichotomous variables for whether the respondent had attended graduate school (**GRADSCH**), or had no college experience at all (**NOCOLL**), came from a household made up of all adults (**NOKID**), or whether the respondent's primary activity was camping (**CAMP**) as opposed to sightseeing. All of these, except **NOCOLL**, were associated with increased likelihood of returning the expenditure survey. Older people were more likely to respond to the expenditure survey. People who traveled longer distances to reach CMS were less likely to respond.

**MLE for spending.** Expenditures for each of the seven items were modeled independently, following the likelihood function in equation 3. For each item, spending within 50 miles of CMS was assumed to be a function of a constant and two variables: number of people for whom the respondent paid (**PAYFOR**) and whether the respondent's primary activity was camping (**CAMP**). In this sample, about two thirds of respondents visited only **CINS**. Slightly more than half (54.8 percent) of the **onsite** contacts were with campers, most of whom were at **CINS** for 3 to 6 days. The rest were on sightseeing trips and were **onsite** for only one day. Nearly half of those who responded to the expenditure survey reported paying only their own expenses, and another 38 percent reported paying for themselves and one other person.

Results for the expenditure modeling were mixed (Table 3). The significance level for  $\rho$  indicates the importance of sample selection for that expenditure item. A nonsignificant coefficient indicates that a simple **Tobit** model could be used to estimate trip expenditures. Sample selection was a significant factor only for privately owned lodging and boat rental expenditures. For the other expenditure items, a **Tobit** model without sample selection was then estimated (Table 4). In general, explanatory power of the variables on expenditures was poor.

Expected values for each expenditure item were calculated from **Tobit** model results (with or without sample selection, as appropriate), and compared to simple sample means (Table 5). As perhaps could be anticipated, expected values for items estimated by a simple **Tobit** model were all slightly higher than the sample mean. However, expected values for **PVTLODGE** and **BOAT** were 40-50 percent less than the sample means. Because **PVTLODGE** is such a large portion of the total trip expenses, total per group per trip expenditures within 50

TABLE 2  
**Probit** Model Results for Returning Expenditure Survey from CINS

Variable	Coefficient	t- value
<b>Constant</b>	-1.1578	-6.07
CAMP	0.3206	3.20
<b>NOKID</b>	0.3406	3.58
AGE	0.0192	5.17
GRADSCH	0.1938	1.92
NOCOLL	-0.2783	-1.86
NMILE	-0.00025	-1.87

**N=812**

Log-likelihood = -527.14

Chi-square = 65.0 1

miles of CINS are 15 percent lower when calculated via **Tobit** models compared to using sample means. After dividing per group per trip expenditures by the sample average of 1.819 people per group, the **Tobit** estimate of per person per trip spending in this economy is **\$50.17**. Of that, \$14.55 is spent on lodging owned by the private sector, and \$14.73 is spent in restaurants. In contrast, the sample mean yields an estimate of **\$59.05** per person per trip, including almost \$25 for privately owned lodging, and \$13.22 at restaurants.

## V. ECONOMIC IMPACT ESTIMATES

The next step examined the effects of these expenditure differences on estimates of economic impacts. For this, the local economy included **Glynn** and **Camden** Counties in Georgia. **IMPLAN** is a **software** and data system for preparing input-output models. **IMPLAN** has been used frequently in estimating the economic effects of natural resource policy decisions, including recreation visitation. **IMPLAN** version 9 1 -F and 1992 base year data were used to estimate impacts. Expenditure items were treated as commodity purchases in various sectors (Table 6). **Spending** on publicly-owned lodging facilities (**PUBLODGE**) was not included, as these fees, usually for campsites, go to general treasuries at either the state or federal level, and thus leak immediately out of the economy. Average per person per trip expenditures were adjusted to 1992 dollars, using deflators included in **IMPLAN** software modules. Impacts per 1,000 nonlocal visitors were estimated.

TABLE 3  
MLE Tobit with Sample Selection Results, for Full Sample of Contacted Visitors  
(N=812)

Variable'	Expenditure Item			
	PVTLODGE	PLJBLDGE	FOOD	RSTRNT
Constant (t-value)	-1.1952 (-6.30)	-1.1718 (-6.14)	-1.1724 (-6.15)	-1.1837 (-6.20)
CAMP	0.2912 (2.90)	0.2969 (2.97)	0.2928 (2.93)	0.2936 (2.94)
<b>NOKID</b>	0.3720 (3.97)	0.3645 (3.82)	0.3699 (3.88)	0.3605 (3.80)
AGE	0.0198 (5.40)	0.0196 (5.25)	0.0193 (5.18)	0.0196 (5.26)
<b>GRADSCH</b>	0.2260 (2.30)	0.2215 (2.20)	0.2247 (2.23)	0.2204 (2.18)
NOCOLL	-0.3078 (-2.14)	-0.2923 (-2.03)	-0.2833 (-1.97)	-0.2941 (-2.02)
<b>NMILE</b>	<b>-0.00022</b> (-1.63)	<b>-0.00026</b> (-1.87)	<b>-0.00022</b> (-1.60)	<b>-0.00022</b> (-1.55)
<b>Constant</b>	51.313 (1.48)	-80.91 (-2.66)	-16.35 (-1.29)	9.0617 (0.61)
<b>PAYFOR</b>	7.4769 (0.80)	2.6051 (0.43)	1.6986 (0.59)	3.9821 (1.16)
<b>CAMP</b>	-66.00 (-3.02)	-1.191 (0.09)	-4.006 (0.61)	-10.72 (-1.33)
<b>Tobit <math>\sigma</math></b>	<b>162.57</b> (12.08)	76.24 (9.82)	49.22 (35.62)	64.58 (26.65)
P	-0.4116 (-1.99)	-0.0987 (-0.27)	<b>-0.0207</b> (-0.08)	<b>-0.1529</b> (-0.61)
Pseudo <b>R<sup>2</sup></b>	<b>.196</b>	<b>.434</b>	<b>.162</b>	<b>.056</b>

TABLE 3 (Continued)  
MLE Tobit with Sample Selection Results,  
for Full Sample of Contacted Visitors

Variable'	Expenditure Item		
	BOATFARE	GAS	s o u v
Constant (t-value)	-1.0872 (-5.79)	-1.1968 (-6.27)	-1.1626 (-6.04)
CAMP	0.2705 (2.72)	0.2914 (2.92)	0.2950 (2.94)
NOKID	0.3808 (4.20)	0.3774 (3.99)	0.3701 (3.88)
AGE	0.0176 (4.83)	0.0193 (5.24)	0.0194 (5.19)
GRADSCH	0.2482 (2.58)	0.2300 (2.30)	0.2254 (2.24)
NOCOLL	-0.3231 (-2.24)	-0.2916 (-2.01)	-0.2906 (-1.87)
NMILE	<b>-0.0025</b> (-1.84)	<b>-0.0018</b> (-1.28)	<b>-0.0026</b> (-1.87)
Constant	2.9686 (0.55)	4.5704 (0.45)	-62.65 (-2.96)
PAYFOR	3.2498 (3.47)	-1.588 (-0.66)	6.1564 (1.41)
CAMP	1.2208 (0.43)	0.5775 (0.12)	-11.74 (-1.36)
Tobit $\sigma$	24.72 (7.74)	38.82 (17.9)	55.29 (9.76)
$\rho$	-0.5771 (-3.13)	<b>-0.2735</b> (-1.13)	<b>0.0369</b> (0.11)
Pseudo $R^2$	<b>.160</b>	<b>.134</b>	<b>.387</b>

'Fii seven estimates are for the probit equation veavor of variables.

TABLE 4  
MLE **Tobit** (without sample selection) Results

Variable	Expenditure Item				
	<b>PUBLODGE</b>	F O O D	RSTRNT	GAS	s o w
constant (t-value)	-84.379 (-4.52)	-16.066 (-2.27)	0.9465 (0.11)	-4.8406 (-0.95)	-62.264 (-5.06)
<b>PAYFOR</b>	0.9066 (0.16)	1.2056 <b>(0.44)</b>	4.309 (1.33)	-0.9848 (-0.49)	5.8016 (1.60)
CAMP	-6.9676 (-0.55)	-4.7483 (-0.82)	-10.864 (-1.50)	0.9250 <b>(.22)</b>	-12.188 (-1.41)
<b>Tobit <math>\sigma</math></b>	78.850 <b>(8.44)</b>	48.468 (15.9)	63.35 (19.0)	36.729 (17.7)	55.612 (8.98)
Pseudo <b>R<sup>2</sup></b>	<b>.560</b>	<b>.467</b>	<b>.329</b>	<b>.384</b>	<b>.559</b>

For the expenditure vector obtained from the sample average, impacts per 1,000 visitors were about \$77,000 in industrial output, \$48,000 in income, and about 1.8 jobs (Table 7). With expenditures calculated as **Tobit** means from the regressions, economic impacts per 1,000 visitors were about \$69,000 in industrial output, \$42,000 in income, and 1.7 jobs. Most of the difference was accounted for in the Hotel and Other Lodging sector. In that sector, impacts **from the Tobit** expenditure vector were about 60 percent of impacts from the sample mean expenditure.

## VI. DISCUSSION

This paper has examined the potential for sample selection bias in visitor expenditure means and in subsequent economic impacts of recreation, **when** visitor expenditure data are collected in the conventional manner. Variables related to self selection of the expenditure respondents may also relate to expenditure levels, although the combined effects may offset one another. Not only should analysts examine demographic variables, including age and education, which were found to be related to selection here, but also trip-related variables that may relate to expenditure patterns including travel distance, length of stay, and group size.

For most of the expenditure items examined here, sample selection was not a significant issue. As well, predicted **Tobit** means for these items were only

TABLE 5  
Expenditures Within 50 miles of CMS: Sample Means and Expected Values

Expenditure Item	Spending per group per trip		Spending per person per trip	
	Sample Mean	Tobit Mean	Sample Mean	Tobit Mean
PVTLODGE <sup>1</sup>	\$45.43	\$26.46	24.98	14.55
<b>PUBLODGE<sup>2</sup></b>	5.23	5.44	2.88	2.99
FOOD <sup>2</sup>	10.31	12.21	5.67	6.71
<b>RSTRNT<sup>2</sup></b>	24.04	26.79	13.22	14.73
<b>GAS<sup>2</sup></b>	9.99	11.81	5.49	6.49
BOATFARE <sup>1</sup>	8.25	4.20	4.54	2.31
<b>SOUV<sup>2</sup></b>	4.17	4.35	2.29	2.39
Total	107.42	91.26	59.05	50.17

<sup>1</sup> Calculated from **Tobit** model, with sample selection.

<sup>2</sup> Calculated from **Tobit** model, without sample selection.

TABLE 6  
Allocation of expenditure items to **IMPLAN** sectors

Expenditure Item	Implan Sector	Sector Description
PVTLODGE	463	Hotels and lodging places
FOOD	450	Food stores
RSTRNT	454	Eating and drinking places
GAS	451	Automobile <b>dealers</b> and service stations
<b>BOATFARE</b>	488	Miscellaneous recreation services
s o w	455	Miscellaneous retail

marginally different than sample means of per group expenditures within the local economy. However, sample selection was significant for the largest expenditure item, privately owned lodging. For that item, the sample mean was over 70 percent higher than the mean **from** the sample selection regression. Resulting industrial output impact estimates in **IMPLAN's** Hotel and Lodging sector were 75 percent higher, and overall impact estimates about 10 percent higher using the uncorrected sample means.

TABLE 7

Economic Impact Estimates, Total Impacts and Selected Sectors		
	Expenditure Vector Source	
	Sample Mean	Tobit Mean
TOTAL <b>INDUSTRIAL</b> OUTPUT (\$ 1000):		
Hotels and other lodging	20.6	12.1
Eating&drinking places	13.5	14.7
Food stores	6.3	7.2
Total, all sectors	76.9	69.0
TOTAL <b>INCOME</b> (\$ 1000):		
Hotels and other lodging	15.5	9.2
<b>Eating/drinking</b> places	6.8	7.4
<b>Food</b> stores	4.5	5.2
<b>Total</b> , all sectors	48.3	42.4
JOBS:		
Hotels and other lodging	<b>.36</b>	<b>.21</b>
Eating/drinking places	<b>.51</b>	<b>.55</b>
Food stores	<b>.26</b>	<b>.30</b>
Total, all sectors	1.78	1.68

These results indicate that a good deal of past research on economic impacts of recreation may be biased due to bias in per trip expenditure vectors, regardless of the economic model used to estimate impacts. Certainly researchers need to consider these issues in designing future studies. In addition, policy makers and policy analysts need to be aware of the pitfalls inherent in this type of survey research.

Although conventional methods here lead to upwardly biased expenditure (and therefore impact) estimates, it is not clear to what degree or in what direction previous efforts may be biased. Undoubtedly, some researchers have weighted their expenditure response sample to reflect certain characteristics of their **onsite** sample. It is not known to what degree such weighting might correct or aggravate selection bias in resulting average expenditure estimates.

Examining the role of sample weighting in resulting expenditure estimates could be a **fruitful** area for future research. A practical question is whether weighted means adequately correct for sample selection and censoring. Future research efforts should ensure that sufficient data are collected in **onsite** contacts

to allow for analysis of selection bias in expenditure survey responses. In addition, assumptions made here about independence across expenditure items or independence of party size and expenditure response could be relaxed.

It is possible that CINS is an atypical recreation site in several respects. Visitors generally are either sightseers who stay one day, or campers who stay at primitive designated campgrounds. In either case, there are very few opportunities to spend money while at CINS or in the surrounding area. In addition, party sizes and lengths of stay are relatively low compared to many other recreation sites. These characteristics could affect the proportion of visitors with positive spending on any item, as well as the number of items purchased.

It is difficult to generalize these results and say that not accounting for sample selection must necessarily cause bias in expenditure means. Rather, this paper has shown that sample selection is an issue the applied researcher must examine, and has provided at least one technique to do so. Accurate estimates of these impacts are important, given the salience of such information for policy making by public agencies and elected officials. Development of a standard set of methods is necessary if researchers hope to develop an impacts transfer framework similar to the benefits transfer concept now being used with net economic surplus in recreation demand studies (Walsh, et al., 1992). Further, when researchers or policy analysts apply impact results **from** a study to another site, or compile results from several individual studies, sample selection or other corrective procedures used in those studies must be evaluated.

## **ENDNOTES**

1. Discussions with Ray Souter, Stan Zamoch, and J.M. Bowker contributed greatly in the preparation of this paper. Also, helpful comments were provided by **Aleigne Amara** and three anonymous reviewers. Bob Leeworthy of the National Oceanic and Atmospheric Administration provided the data.

2. Sample selection can also lead to biased estimates of party size.. However, the variation in this variable is usually small, as most parties are made up of 4 or fewer people. For the purposes of this paper, it is assumed that sample selection has no measurable effect on estimates of party size.

3. In order to be consistent with the paper's focus on sample selection, including income would have entailed incorporating a second selection equation. There is no clear theoretical reason why income should be related to returning the expenditure survey. Although total expenditure on recreation is a normal good and can thus be theoretically linked to income, making the same claim about the amount spent on one given trip in a small geographic area stretches economic

theory a bit too far. One reviewer expressed concern that excluding income admits the possibility of misspecification bias. This point can be debated. In this case model simplicity and an unclear conceptual mandate outweighed any small concern for misspecification.

## REFERENCES

- Alward**, G. S. and E. M. Lofting. "Opportunities for Analyzing the Economic Impacts of Recreation and Tourism Expenditures Using IMPLAN." 'Contributed paper, Annual meeting of the Regional Science Association. Philadelphia, Pennsylvania, 1985.
- Alward**, G. S., H. C. Davis, **K.** A. Despot&is, and E. M. Lofting. "Regional Non-survey input-output analysis with **IMPLAN.**" Contributed paper, Southern Regional Science Association Annual meeting. Washington, DC, May 1985.
- Bergstrom, J. C., H. **K.** Cordell, A. E. Watson and G. A. Ashley. "Economic Impacts of State Parks on State Economies in the 'South.'" *Southern Journal of Agricultural Economics*. 22 (1990). 69-77.
- Cochran**, W.G. *Sampling Techniques*. 3rd. edition. New York: John Wiley and Sons, 1977.
- Cordell, H.K., J.C. Bergstrom, G.A. Ashley, and J. **Karish**. "Economic effects of river recreation on local economies." *Water Resources Bulletin*. 26 (1990). **53-60**.
- Cordell, H.K., J.C. Bergstrom, and A.E. Watson. "Economic growth and Interdependence Effects of State Park Visitation in State and Local Economies.\*" *Journal of Leisure Research*. 24 (1992). 253-268.
- Dawson, S.A., D.J. Blahna, and J.E. Keith. "Expected and actual regional economic impacts of Great Basin National Park." *Journal of Park and Recreation Administration*. 11 (1993). 4559.
- Dillman**, D.A. *Mail and Telephone Surveys*. New York: John Wiley and Sons, 1978.
- Dobbs, LM. "Adjusting for sample selection bias in the individual travel cost method." *Journal of Agricultural Economics*. **44** (1993). 335-342.
- English, D.B.K. and J.C. Bergstrom. "The Conceptual Links Between Recreation Site Development and Regional Economic Impacts.\*" *Journal of Regional science*. 34 (1994). 599-6 11.
- Greene, W.H. *Limdep 7.0: User's Manual*. Bellport, NY: Econometric **Software**, Inc., 1995.

- Heckman, J.J.** "Sample selection bias as a specification error." *Econometrica*. 47 (1979). 153-161.
- Leitch, J. A., and F. L. **Leistritz**. "Techniques for Assessing the Secondary Economic Impacts of Recreation and Tourism." *In Assessing the Economic Impacts of Recreation and Tourism*. ed. D. B. Propst. Asheville, NC: Southeastern Forest Experiment Station, 1985.23-27.
- Jackson, R. S., D. J. **Stynes**, D. B. Propst, and L. E. Siverts. "Economic Impact Analysis as a Tool in Recreation Program Evaluation." Instruction Report R-92-1. Vicksburg, MS: U.S. Army Engineer Waterways Experiment Station, 1992.
- Johnson, R. L. and E. Moore. "Tourism Impact Estimation." *Annals of Tourism Research*. 20 (1993). 279-288.
- Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs No. 3. Cambridge: Cambridge University Press, 1983.
- Nawata, K.** "Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator.\*" *Economics Letters*. 45 (1994). 33-40.
- Propst, D. B., D. J. Stynes, J. H. Lee, and R. S. Jackson.** "Development of Spending Profiles for Recreation Visitors to Corps of Engineers Projects." Technical Report R-92-4, U.S. Army Engineer Waterway Experiment Station, Vicksburg, MS. 1992.
- Schorr, M. S., J. Sah, D. F. Schreiner, M. R. **Meador**, and L. G. Hill. "Regional Economic Impact of the Lake Texoma (Oklahoma-Texas) Striped Bass Fishery." *Fisheries*. 20 (1995). 14-18.
- Shaw, D. "**Onsite** samples regression: problems of non-negative integers, truncation, and endogenous stratification.\*" *Journal of Econometrics*. 37 (1988). 211-223.
- Smith, V. K. "Selection and Recreation Demand." *American Journal of Agricultural Economics*. 70 (1988). 29-36.
- Stevens, B. and A. Rose. "Regional Input-Output Methods for Tourism Impact Analysis." *In Assessing the Economic Impacts of Recreation and Tourism*. ed. D. B. Propst. Asheville, NC: Southeastern Forest Experiment Station, 1985. 16-22.
- USDA Forest Service. "Forest Service Program for Forest and Range Resources: Long Term Strategic Plan." Draft 1995 **RPA** Program. Washington, DC: U.S. Department of Agriculture, Forest Service, 1995.
- Walsh, R. G., D. M. Johnson, and J. R. **McKean**. "Benefit Transfer of Outdoor Recreation Demand Studies, 1968-1988." *Water Resources Research*. 1992. **28(3)**. 707-713.

Willis, K., and G. Garrod. "Valuing Open Access Recreation on Inland Waterways: On-site Recreation Surveys and Selection Effects." *Regional Studies*. 25 (1991). 51 1-524.

**Winship, C.**, and R. D. Mare. "Models for Sample Selection Bias." *Annual Review of Sociology*. 18 (1992). 327-350.