JOURNAL OF REGIONAL SCIENCE, VOL. 40, NO. 3, 2000, pp. 523–539

CALCULATING CONFIDENCE INTERVALS FOR REGIONAL ECONOMIC IMPACTS OF RECREATION BY BOOTSTRAPPING VISITOR EXPENDITURES*

Donald B. K. English
USDA–Forest Service, Southern Research Station, 320 Green Street, Athens, GA 30602. E-mail:denglish@fs.fed.us

ABSTRACT. In this paper I use bootstrap procedures to develop confidence intervals for estimates of total industrial output generated per thousand tourist visits. Mean expenditures from replicated visitor expenditure data included weights to correct for response bias. Impacts were estimated with IMPLAN. Ninety percent interval endpoints were 6 to 16 percent above or below the original sample's point estimate depending on the calculation method. Due to the linearity of input-output a shortcut method that estimates confidence interval endpoints from the distribution of mean expenditure profiles yields nearly identical results.

1. INTRODUCTION

The National Environmental Policy Act (42 U.S.C. 1500) mandates that federal agencies consider the economic consequences of their management actions. These consequences include both economic valuation estimates as well as local or regional economic impacts. Many federal agencies, including the USDA Forest Service and the USDC National Oceanic and Atmospheric Administration, include both valuation and impact measures as appropriate evaluation criteria for planning, formulating resource policy, and assessing environmental damages (Forest Service, 1995; NOAA, 1995).

Measuring economic welfare and regional impacts of recreation visitation are among the issues that are most contentious and difficult to quantify in resource allocation decisions. Estimates of these economic measures are usually based on data obtained from surveys of visitors to public recreation sites. Travel cost or willingness-to-pay questions typically provide information for estimating economic values from demand functions (Smith, 1993). Means of visitors' expenditure profiles per trip are the source for final-demand changes used in economic impact evaluations (Johnson and Moore, 1993; Douglas and Harpman, 1995; Bergstrom et al., 1996).

*V. R. Leeworthy, National Oceanic and Atmospheric Administration, provided the visitor data for this paper. Discussions with Steven Deller, Dave Marcouiller, Ron Schaeffer, and J. M. Bowker were helpful. The comments and suggestions of three anonymous reviewers are also gratefully acknowledged.

Received January 1999; revised September 1999; accepted January 2000.
© Blackwell Publishers 2000.
Blackwell Publishers, 350 Main Street, Malden, MA 02148, USA and 108 Cowley Road, Oxford, OX4 1JF, UK.

The estimators for values and impacts are both random variables because of the process that generates them. The true distributions of those estimators are not always easy to determine or describe. Analysts who want more than just point estimates of the values or impacts per visit must rely on some other method to determine the statistical accuracy of the economic measures. Some work has been done on evaluating the distribution of valuation estimates but no analogous research has been done for economic impacts of recreation. In this paper I use a resampling technique known as bootstrapping to illustrate a method by which confidence intervals for economic impacts of recreation visitation may be developed.

2. TYPICAL PROCESS

In a typical study to estimate the economic value or regional economic impact of visitation to a recreation site, a (possibly stratified) random sample of n visits to that site are drawn. Intercept surveys occur as the recreation visit ends. The person making the visit is asked the questions needed for valuation estimation including the number of annual visits to the site, travel distance and time, and some "demand shifters," such as gender, income, or educational attainment. At that time, or in a mailed follow-up survey, the person provides information about the amount of money spent on a set of k expenditure items for the visit. Invariably, not all of those contacted respond, yielding m_v ($< n$) usable responses for valuation and m_e ($< n$) expenditure responses.

In welfare studies, individual trip price is computed from monetizing reported travel distance and travel time. This price and some number of demand shifters are regressed on annual visitation rates. Assumptions about functional form and error distribution determine the regression structure. For example, count data models reflect obvious restrictions on trip-taking behavior. Average per-trip consumer surplus is a function of the estimated coefficient on the price term.

For an impact study, let $\mathbf{E}$ be the $m_e \times k$ matrix of expenditure data. To account for sample stratification and to correct for nonresponse bias (see, for example, Leeworthy et al., 2000) a ($m_e \times 1$) vector of weights $\mathbf{W}$ is constructed. The estimate of average expenditures on each item $\mathbf{X}$ is

$$\mathbf{X} = \left(\frac{1}{\sum_{i=1}^{m} W_i} \right) \mathbf{E}'\mathbf{W}$$

To facilitate visitor response, expenditure items on surveys conform to the types of goods and services visitors typically purchase. However, models of regional economies are most often based on industrial sectors. It is seldom the case that there is a one-to-one mapping of survey items to industries. As a result, the $k \times 1$ vector of mean expenditures must be 'bridged' onto the j industrial

© Blackwell Publishers 2000.

sectors in the economic model. Let **B** be the $k \times j$ bridging matrix that maps the expenditure vector onto the industrial sectors. The vector **D** that describes the demand shock to the economy from the average recreation visit is

$$\mathbf{D} = \mathbf{B}'\mathbf{X}$$

Input-output models are widely applied and mathematically straightforward models of regional economies (Miller and Blair, 1985). The demand vector for one visit may not have a measurable effect on a regional economy, so **D** is often scaled upward to represent a thousand visits. Given the standard **A** matrix of technical coefficients, the impact **P** of the vector **D** on the economy is

$$\mathbf{P} = (\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{D}$$

It is this vector or often simply the sum of its j elements, that is, the total economic impact, that is of primary interest.

3. VARIATION IN VALUATION AND IMPACT ESTIMATES

The two economic measures have analogous sources of variation. Practitioners treat these sources of variation in roughly the same way. Estimates of both impacts and valuation are based on information provided by a random sample of people contacted during a recreation visit. Characteristics of the intercept sample (such as mean age or percent female) are random variables. As a result weights that correct for stratification or selection bias are also random variables. However, in both types of studies the intercept sample is usually treated as representative. Thus, variation that may come from differences in the intercept sample is ignored. Another such source is in measurement error that may exist in information collected from visitors. Respondents may round their number of trips, miles traveled, or dollars spent to some convenient multiple of 5, 10, or 100. Others simply may not remember accurately.

A number of other sources of variation are assumed away in the standard estimation process for both economic measures. In valuation work the economic construct of the individual's trip price is computed by monetizing travel time and distance. The time cost portion of trip price is assumed to equal a fixed percentage of the person's hourly wage rate multiplied by the number of hours traveled. Monetary travel costs are usually computed as a constant cost per mile multiplied by the number of miles traveled. Undoubtedly, the formula that correctly computes the value of travel time and out-of-pocket costs per mile of travel will not be the same for all individuals who visit the recreation site. However, the true distributions of time value and out-of-pocket travel costs across the population of individuals are not known. One formula is assumed to be known and to be equally applicable to all visitors. As a result, this source of variability is not included in valuation studies. In impact studies, the bridging matrix **B** serves the same function as the monetizing formulae in valuation work. The construct of a final-demand vector is created from reported spending on commodities. The transformation is accomplished by

© Blackwell Publishers 2000.

applying a set of coefficients that are assumed to be known and constant across all individuals.

The model of individual behavior defined by the regression is at the heart of valuation work. Assumptions about the error distribution, explanatory variables, and an appropriate functional form determine the model's structure. The true structure is not known, nor is how it might vary across individuals. Final results about per-person or per-trip surplus are strongly influenced by what model is used (Adamowicz, Fletcher, and Graham-Tomasi, 1989). In impact analysis the regional economic model is the central element. A matrix of technical coefficients is assumed to accurately capture the interlinked behavior of industries. It is assumed that these coefficients apply equally to all firms in any given industrial sector. Again, the true distribution of these coefficients across the firms in the region is unknown and is not included in the impact calculations.

Clearly, the estimators of both economic values and impacts of recreation visitation are random variables. Due to the many sources of variation that could be included in the estimation process, the true probability distributions for the estimators may be extremely difficult to determine. However, standard practices simplify the situation. Reported estimates are contingent on a number of assumptions, and these assumptions remove a number of the sources of variability. The primary source of variation that remains is in individuals' reported expenditures and annual visit rates. For valuation, annual visitation is often modeled using some form of count-data model, which assumes a Poisson or negative binomial error term. Per-visit spending on individual items often includes a large proportion of zeros and a few observations with extremely large expenditures (Leeworthy et al., 2000).

Point estimates of average benefits or impacts per trip derived from a single sample of visitors may not be sufficient information for making optimal resource allocation decisions. Understanding and accounting for the variability of such measures may also be necessary. For example, in a benefit-cost framework the analyst may need to know how likely it is that benefits will exceed project costs (Adamowicz, Fletcher, and Graham-Tomasi, 1989), in addition to knowing whether the expectation of benefits will exceed expected costs. Such information may be especially important if reversing the decision is costly. Determining that likelihood requires knowledge about the estimator's distribution.

The Central Limit Theorem indicates that the mean from a random sample will be asymptotically normally distributed regardless of the underlying distribution from which the sample is drawn as so long as that distribution has a finite variance. That is, over an infinite number of repeated samples from the same population, the average impact per visit or the average surplus per visitor will have approximately a normal distribution. Unfortunately, neither the mean nor the variance of the distribution are known, nor is it clear how quickly the convergence to normality occurs. The problem is how to estimate the distribution of the estimator from only one sample of on-site intercept surveys.

© Blackwell Publishers 2000.

4. BOOTSTRAPPING TO THE RESCUE

The bootstrap method (Efron and Tibshirani, 1993) allows an analyst to obtain an approximation of the distribution of an estimator in the absence of a priori information about the true distribution of the estimator or the original data. From a sample of size n, a large number of new data sets are generated by drawing, n observations with replacement from the original sample. The estimator is calculated for each new data set. The resulting empirical distribution of estimator values is used to approximate its true distribution. Several methods are available for developing bootstrap confidence intervals including the normal approximation, percentile, and bootstrap-t.

The normal approximation begins with θ, the value of the estimator from the original sample. Let σ^* be the standard deviation of the bootstrap distribution of results of estimator values. Given the Z value from a standard normal distribution that gives the desired confidence level α, the confidence interval for the normal approximation is defined as $\theta \pm z_{\alpha/2}\,\sigma^*$.

Percentile methods for confidence intervals begin by ordering the bootstrap results from lowest to highest. Let $\theta^*(b)$ be the impact result for the bth replicate, and $F(\theta^*)$ be the empirical cumulative density function (c.d.f.) for the bootstrap results. For a desired confidence level α the values of $F(\theta^*)$ at the $\alpha/2$ and $100 - (\alpha/2)$ percentiles are the lower and upper confidence bounds, respectively. However, $F(\theta^*)$ may not be centered on θ. The bias-corrected percentile (BCP) method corrects for such a condition.

Let Φ represent the standard normal c.d.f., and let $z_0 = \Phi^{-1}\{\Pr(\theta^* \leq \theta)\}$. That is, z_0 locates the original sample estimate within $F(\theta^*)$. For example, if 67 percent of the bootstrap distribution is less than or equal to θ then $z_0 = 0.44$. For desired confidence level α, the lower bound on the confidence interval is the value of $F(\theta^*)$ at the $[\{\Phi(2z_0 + z_{\alpha/2})\} \times 100]$ percentile. The upper bound is the value of $F(\theta^*)$ at the $[\{\Phi(2z_0 + z_{1-\alpha/2})\} \times 100]$ percentile. Continuing with the example, if a 95 percent confidence interval is desired the lower bound is $\Phi[2(0.44) - 1.96]$ or the 14.01 percentile of the empirical distribution, and the upper bound is $\Phi[2(0.44) + 1.96]$ or the 99.77 percentile.

Like percentile methods, the bootstrap-t method also computes intervals without relying on a normal theory assumption. Compute

$$Z^*(b) = \frac{\left[\theta^*(b) - \theta\right]}{\sigma^*(b)}$$

where $\sigma^*(b)$ is the estimated standard error for the bth bootstrap replicate. A percentile method is used to determine the values for a t-table (t^*) suitable to the data at hand. That is, for 1,000 replicates, the t^*-value for the 5 percentage point is the 50th smallest value of the $Z^*(b)$s and the t^*-value for the 95 percentage point is the 950th smallest value. Bootstrap-t confidence intervals are defined by

© Blackwell Publishers 2000.

$$[\theta - (t^{*(1-\alpha)})\,\sigma^*, \theta + (t^{*(1-\alpha)})\,\sigma^*]$$

The formula for computing $Z^*(b)$ requires a bootstrap estimate of standard error for each bootstrap sample, usually necessitating nested bootstrap sampling (Efron and Tibshirani, 1993).

5. VARIABILITY RESEARCH IN RECREATION VALUATION

Examining the variability in welfare or valuation estimates of recreation via resampling has been addressed by several research efforts. Most have followed a modified Krinsky-Robb procedure to accomplish the resampling. For example, Creel and Loomis (1991) drew a random sample of 8,000 parameters from an assumed multivariate normal distribution whose mean vector and covariance matrix were defined by the parameters estimated in a travel-cost demand equation. Ninety-percent confidence intervals for welfare measures were defined by a percentile method, wherein the results were ordered and 5 percent of observations were removed from each tail.

Adamowicz, Fletcher, and Graham-Tomasi (1989) created new dependent variables from the observed data matrix and random draws from a normal error term with mean zero and variance determined by the error in the regression for several functional forms of travel-cost demand. Regressions on the new dependent variables yielded a new estimate of coefficients and ultimately welfare measures. Repeating this process 5,000 times for each functional form provided a distribution of welfare measures from which means and standard deviations were reported.

Kling and Sexton (1990) followed a process similar to Adamowicz, Fletcher, and Graham-Tomasi, but they drew a bootstrap sample from the empirical regression error distribution, rather than from an assumed normal error distribution. In addition, they eliminated bootstrap results wherein willingness to pay was less than zero or greater than total income. For each of 16 data sets, one hundred bootstrap trials were generated from which coefficients of variation were reported. Confidence intervals were calculated as if the bootstrap trial results were normally distributed.

Yen and Adamowicz (1993) combined the Krinsky-Robb procedure used by Creel and Loomis with the theoretical restrictions-to-consumer-surplus results of Kling and Sexton. For each of several models, 10,000 vectors of parameters were drawn. Ninety-percent confidence intervals were reported, presumably calculated via a percentile method because the intervals are not symmetric about the mean of the simulation results.

Resampling has also been used to assess the variability of welfare estimates in some contingent valuation studies. Park, Loomis, and Creel (1991) and Souter and Bowker (1994) used a Krinsky-Robb approach. In both applications confidence intervals were based on 1,000 replicates and a percentile method for determining interval endpoints. Cooper (1994) used bootstrapping as well as

© Blackwell Publishers 2000.

Krinsky-Robb and analytic approaches to evaluate confidence intervals for welfare estimates from dichotomous choice CVM.

However, variability of results has not been addressed in empirical research on the economic impacts of outdoor recreation or resource-based tourism. Only point estimates of the impacts per 1,000 visitors are reported, either for total impacts or the vector of impacts per economic sector. In some instances, means for the vector of visitor expenditures are available (Uysal, Pomeroy, and Potts, 1992; Johnson and Moore, 1993). In some studies, visitors have been segmented into groups that are expected to have homogeneous spending patterns, and the means per group are indicated (Propst et al., 1992).

Why has variability in impacts not been addressed? Could it be that regional scientists recognize that interpersonal variation in visitor expenditures is not the greatest source of uncertainty in estimates of economic impacts? There is little doubt that variation in the bridging or technical coefficients matrices exists across economic actors. Further these sources of variation could easily dwarf variation in expenditures. An analogous claim (i.e., that interpersonal variation in the trip demand and computed price are not the sole or even largest source of uncertainty) could be made for valuation. Yet not incorporating all sources of variation has not stopped work on valuation research. In the research reviewed above the only source of variation has been in interpersonal differences in the variables included in the regression equation.

The bootstrapping procedure for economic impacts requires developing B bootstrapped data sets and following the standard process through to an estimate of impacts $\mathbf{P}$ for each. Confidence intervals may be calculated from these results. Unfortunately, it is time consuming to compute 1,000 or so impact estimates with programs like IMPLAN.

In the remainder of this paper I present an empirical example of developing confidence intervals for economic impacts of recreation. In addition, the policy implications of the magnitude of the intervals relative to the point estimate are discussed. A critical issue is the cost involved in developing confidence intervals. No matter how straightforward the procedure, confidence intervals will not be computed in most practical applications if it is expensive. Generally, the time cost in developing confidence intervals is directly related to the number of bootstrap replicates. Consequently, some attention is devoted to the stability of the interval estimate relative to the number of bootstrap replicates.

In the following section I describe the empirical data used in this paper. In Section 7 I describe methods used to determine nonresponse weights, bootstrap replicates, and confidence intervals. I present and compare the results of the different methods in Section 8. In the final section I provide a discussion and conclusion.

6. DATA

The data were collected though an interagency agreement between the National Oceanic and Atmospheric Administration (NOAA), the USDA–Forest

© Blackwell Publishers 2000.

Service, and the University of Georgia's Department of Agricultural and Applied Economics. The study was designed to collect data necessary to estimate economic impacts and values for recreational visits to the Florida Keys. Separate samples were developed for winter and summer because of seasonal differences in visitation, weather, and resource uses. The summer survey period was July and August of 1995. In this paper I make use of the summer-season data only.

An on-site random-intercept survey, stratified by mode of travel (air, auto, cruise ship) provides information for demographic profiles of the visitor population. Detailed visitor profiles may be found in Leeworthy and Wiley (1996). Each person contacted during the on-site exit interview was given an expenditure questionnaire to fill out and mail back. Expenditure information for 50 different trip-related expenditure items was obtained via the mail survey. There were five general types of expenditures: lodging (7 items), food (3 items), transportation (9 items), activities (22 items), and miscellaneous (9 items). For each item, respondents were asked how much they spent in total on the trip and specifically how much they spent in the three-county south Florida area. Following Dillman (1978), reminder postcards and second questionnaire mailings were made at two-week intervals as needed. Of the 1,334 contacts, 505 people (37.86 percent) responded to the mailback expenditure survey.

7. METHODS

Nonresponse Bias Weights

Several tests for nonresponse bias were conducted because of the relatively low response rate. The testing process is described in detail in Leeworthy (1996). Several variables were related to both response rates to the expenditure survey and to the amount of reported spending in south Florida. For example, foreign visitors were less likely to respond than domestic (U.S.) visitors, but foreign visitors also spend more money per trip. Other significant variables include race, age, and income. Observation-specific weights were constructed as the product of a stratum weight (to account for the sample design), and a nonresponse bias weight for a demographic category (defined by combinations of race, age group, income class, and residence).

Bootstrap Process

The expenditure subsample provided the multivariate data from which the bootstraps were developed. Using the random-number–generating procedure in the SAS program's UNIFORM function, 1,000 bootstrap samples equal in size to the original expenditure sample (505 observations) were generated.[1] It was necessary to recalculate the response-bias weights for each bootstrap replicate

[1]Original sample observations were numbered from 1 to 505. Next, 505,000 random variates in the 0–1 interval were generated from the RANUNI function, using a seed generated from the SAS clock. The observation number corresponding to each variate was determined by multiplying the variate value by 505, adding 0.4999, and rounding to the nearest integer.

© Blackwell Publishers 2000.

because of concerns about nonresponse bias in the original expenditure sample. The process was similar to that used in the original sample. That is, the proportion of cases in the replicate in each demographic categories was calculated for each replicate. The nonresponse weight for the category is the proportion of replicate cases in the category divided by the proportion of cases from the on-site sample in that category. The weighted average expenditure vector **X** for each item was obtained for each bootstrap replicate. The bridging matrix **B** used to calculate final-demand vectors **D** is described in Leeworthy et al. (1996). Economic impacts per 1,000 visitors were estimated for each of the 1,000 bootstrap samples using IMPLAN and 1993 structural data. The primary economic measure of interest is Total Industrial Output (direct + indirect + induced).

Normality Tests

The Komolgorov-Smirnov two-sample criterion D test (Blalock, 1979; Mood, Graybill, and Boes, 1974) is used to compare the bootstrap distributions to a normal distribution with the same mean and variance for a given variable. The test statistic measures the maximum deviation of the empirical distribution function to the statistical distribution:

$$D = \mathrm{Max}_j \mid F_1 \times F_2 \mid$$

where F_1 indicates the empirical distribution function for the bootstrap sample, and F_2 is the statistical distribution function given the same mean and variance as in F_1. If the value of D is greater than some critical value the null hypothesis that the observed distribution is drawn from the distribution function F_2 is rejected. Critical values for D for several sample sizes may be found in Beyer (1966).

A generalized-distance test (Johnson and Wichern, 1992, p. 160) is used to test for joint normality of the vector of expenditure items. If the parent population is multivariate normal then the squared generalized distances

$$d_j^2 = (x_j - x)\, S^{-1}\, (x_j - x)$$

should behave like a chi-square random variable. The test involves ordering the distances and then plotting against $\chi_p^2\ [(j - 1/2)/n]$, where p is the number of variables, n is the number of observations, and j is the number of the observation ($j = 1,2, \ldots n$). If the distribution is multivariate normal then the resulting plot should approximate a straight line. In developing the bootstrap-t confidence intervals, standard errors are estimated from a random draw with replacement of 50 of the 1,000 bootstrapped results. For each bootstrap method, intervals are computed for the 90th, 95th, and 99th confidence levels.

© Blackwell Publishers 2000.

8. RESULTS

Bootstrapped Impact Results

The estimate of economic impacts per 1,000 visitors from the original expenditure sample is 1,020.2 thousand dollars. The mean of the 1,000 bootstrapped results is 980.03 thousand dollars with a standard deviation of 66.264 thousand dollars. Almost three-quarters (74.6 percent) of the bootstrapped results are less than the original sample estimate, so the z_0 value for bias correction was 0.6464. Confidence intervals results for the three methods are presented in Table 1.

The 90-percent confidence interval for impacts per 1,000 visitors under the normal approximation span the range between $911,200 and $1,129,200, or about plus or minus 10.7 percent of the sample estimate value. Bounds for the 99 percent level were at $849,500 and $1,190,900. To evaluate the appropriateness of this method, the empirical distribution of bootstrap results was tested for normality, using the Komolgorov-Smirnov (K-S) D test. For the bootstrap sample of total impact results the value of the D statistic was 0.0251. With n = 1,000, the null hypothesis that the empirical distribution is normal cannot be rejected at the $p < 0.05$ level.

The bootstrap distribution was centered somewhat below the estimate obtained from the original sample. As a result, percentiles for the intervals are shifted upward. For example, the lower bound for the 90-percent confidence interval under the BCP method is at the 36.2th percentile of the bootstrap distribution and the 10th percentile for the 99-percent confidence level. The upper bound for the 90-percent confidence level is at about the 99th percentile

TABLE 1: Confidence Intervals for Bootstrapped Impacts Per 1,000 Visitors (Interval Bounds are in Thousands of 1993 Dollars)[a]

	Confidence Level		
Method/Statistic	90 percent	95 percent	99 percent
Normal Approximation			
Lower Bound	911.2	890.3	849.5
Upper Bound	1,129.2	1,150.1	1,190.9
Bias-Corrected Percentile			
Percentile value for Lower Bound	36.2	25.2	10.0
Percentile value for Upper Bound	99.4	99.9	100.0
Lower Bound	955.1	935.1	895.0
Upper Bound	1,188.1	1,197.4	1,221.0
Bootstrap-t			
t-value for Lower Bound	–2.253	–2.692	–3.198
t-value for Upper Bound	1.137	1.519	2.143
Lower Bound	870.9	841.8	808.3
Upper Bound	1,095.5	1,120.9	1,162.2

[a] Point estimate of impacts per 1,000 visits from the original sample was 1,020.2 thousand dollars.

© Blackwell Publishers 2000.

and the upper bound for the 99-percent confidence level is the largest value in the bootstrap sample. In turn, the confidence interval bounds for this method are above those for the normal intervals. Confidence interval bounds for the 90-percent confidence level are 955,100 to 1,188,100 dollars and for the 99-percent level are 895,000 to 1,221,000 dollars.

As with the percentile method, confidence bounds for the bootstrap-*t* method are also asymmetric about the sample mean. In contrast to the BCP method, this method estimates confidence intervals with both endpoints shifted downward (closer to zero) compared to either of the other two methods. The 90-percent confidence level bounds are at 870,900 and 1,095,500 dollars; the 99-percent interval covers the range from 808,300 to 1,162,200 dollars.

Developing the set of 1,000 bootstrapped impacts estimates took more than 120 hours of programming and computer analysis time. Each impact result had to be individually calculated with IMPLAN. That amount of spare time is not available for many resource planning efforts. Consequently, some means of reducing the cost of developing confidence intervals is needed. The most obvious method is to reduce the number of bootstrap replicates. Efron and Tibshirani (1993) indicated that 1,000 replicates is a reasonable minimum for either BCP or BT methods. However, a normal-approximation method may need substantially fewer replicates to obtain a stable estimate of the standard error.

The bootstrap estimate of the standard error of total impacts was calculated for the first 300 replicates. The next replicate were added to the sample, and the standard error was recalculated. Figure 1 shows the relation between the number of bootstrap replicates and the estimate of the standard error. For these data it appears that over 500 and perhaps as many as 800 replicates are needed to obtain a stable estimate of the standard error.

Another possible shortcut is to develop bootstrapped confidence intervals for per-person per-trip expenditures. Impact estimates only need to be developed for the expenditure vectors that represent the endpoints of the confidence intervals. Impact estimates are straightforward transforms of the expenditure vector so such a shortcut might lead to confidence intervals that are close to those identified above.

Expenditure Vector Intervals

Marginal distributions for expenditure means were tested for normality using the K-S test. Distributions for all 50 spending items as well as for total spending were examined. The null hypothesis that the empirical distribution of bootstrap means was normal was rejected at the $p < 0.01$ percent level for nineteen of the expenditure items. It was rejected at the $p < 0.05$ percent level for seven other items. Pursuit of a multivariate confidence region was abandoned given the variety of distributions for individual expenditure items. Fortunately, the null hypothesis of normality could not be rejected for total expenditures. Confidence intervals for total expenditures were estimated from the bootstrap data.

© Blackwell Publishers 2000.

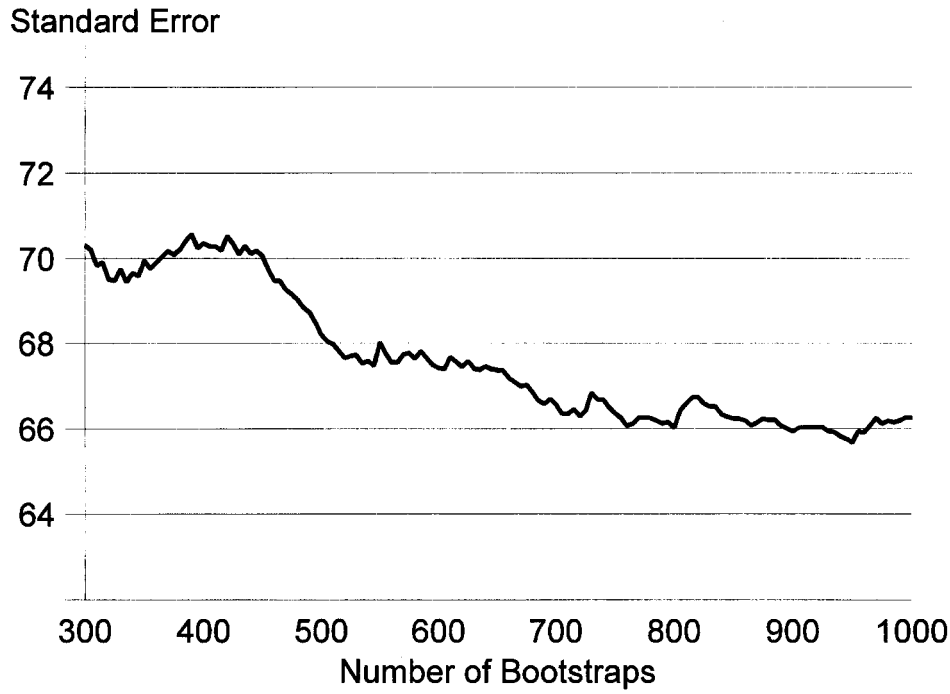

FIGURE 1: Relationship of Bootstrap Standard Error Estimate to Number of Bootstrap Replicates.

Total expenditures in south Florida per 1,000 visitors was estimated at 620,561 dollars from the original sample. The mean of the bootstrapped replicates was 594,583 dollars with a standard deviation of 39,983 dollars. A normal approximation method thus yielded 90-percent confidence intervals ranging from 554,800 to 686,300 dollars (Table 2).

About 76 percent of the bootstrapped expenditure totals were below the original sample estimate. As a result, endpoints for BCP intervals are shifted upward compared to the normal method. The 90-percent interval bounds were 585,200 and 725,000 dollars. Midpoint of the bootstrap distribution was sufficiently high that BCP upper bounds used only the two highest values from the bootstrap sample. As with the impact analysis, bootstrap-*t* intervals were lower than the normal intervals and slightly wider. Bounds for 90-percent intervals are 526,900 and 662,600 dollars.

Confidence bounds on impacts were developed from total expenditure interval bounds by allocating total expenditures across all items in the same relative proportion as for the original sample, then calculating final demand and impacts as described above. As expected from the linear nature of input-output analysis, impact results from the normal intervals for total expenditures were basically identical to (within $1,000 of) the bootstrapped impact results

© Blackwell Publishers 2000.

TABLE 2: Confidence Intervals for Bootstrapped Spending in South Florida Per 1,000 Visitors (Interval Bounds are in Thousands of 1993 Dollars)[a]

	Confidence Level		
Method/Statistic	90 percent	95 percent	99 percent
Normal Approximation			
Lower Bound	554.8	542.2	517.6
Upper Bound	686.3	698.9	723.5
Bias-Corrected Percentile			
Percentile value for Lower Bound	41.3	29.6	12.5
Percentile value for Upper Bound	99.9	100.0	100.0
Lower Bound	585.2	572.2	549.1
Upper Bound	725.0	742.4	742.4
Bootstrap-*t*			
t-value for Lower Bound	–2.342	–2.646	–3.203
t-value for Upper Bound	1.053	1.414	2.034
Lower Bound	526.9	514.8	492.5
Upper Bound	662.6	677.1	701.9

[a]Point estimate of total expenditures per 1,000 visits from the original sample was 620,561 dollars.

(Table 3). Differences are due to rounding errors. Lower bounds for BCP expenditure intervals were between 5,000 and 7,000 dollars higher than in the bootstrapped impact results. Upper bounds were very close to the bootstrapped results for the 90- and 99-percent intervals, but over 23,000 dollars higher for the 95-percent interval. Impact interval endpoints generated from expenditure interval bounds under the bootstrap-t method were also reasonably close to the bootstrapped impact results, differing by not more than 5,000 dollars across the three lower bounds, and by not more than 9,000 dollars for upper bounds.

9. DISCUSSION

Developing confidence intervals from bootstrapped impact results is not overly difficult, although tedious and time consuming. Two different software packages (SAS and IMPLAN) were used in this study. Programming and analysis time combined took more than the equivalent of three work weeks. Software advances may permit automation of IMPLAN analyses, which will greatly reduce the time needed to generate bootstrap intervals. Fortunately, a shortcut method appears to give comparable results. Bootstrap confidence intervals for total spending per trip in the target economy yielded interval bounds on total impacts that differed from those developed directly from bootstrapped impact results by less than one percent. Thus, this shortcut appears to have promise as a "quick and dirty" method for generating rough estimates of confidence intervals for total impacts. However, because the shortcut does not allow any variation in the proportion of spending per sector, confidence intervals for specific sectors generated by this shortcut may be artificially low.

© Blackwell Publishers 2000.

TABLE 3: Confidence Intervals for Impacts in South Florida Per 1,000 Visitors, from Total Expenditure Interval Bounds (Interval Bounds are in Thousands of 1993 Dollars)

	Confidence Level		
Method/Statistic	90 percent	95 percent	99 percent
Normal Approximation			
Lower Bound	912.2	891.3	850.8
Upper Bound	1,128.3	1,149.3	1,189.6
Bias-Corrected Percentile			
Lower Bound	962.0	940.7	902.8
Upper Bound	1,191.9	1,220.6	1,220.6
Bootstrap-*t*			
Lower Bound	866.2	846.3	809.6
Upper Bound	1,089.4	1,113.2	1,153.9

Three methods were used to develop confidence intervals for total impacts from the bootstrapped sample of impact results. Normal approximation results were easiest to compute, but required a specific distributional assumption (justified in this case) about the population of estimators. Another drawback was that only information about the variation in the bootstrap replicate sample was used. Other information, such as the location of the bootstrap sample relative to the original point estimate, was not. The other two methods use somewhat more of the information contained in the bootstrap sample.

Results from the BCP methods are somewhat problematic. Upper endpoints for confidence intervals were constrained by the top end of the bootstrap distribution because of the divergence between the original sample estimate and the midpoint of the bootstrap distributions. For the impacts results, any confidence level above 95 percent have had the largest bootstrap result as the upper bound. For the expenditure results, and confidence level above about 92 percent would have had the same upper bound. A greater number of bootstrap replicates may alleviate this problem.

Intervals obtained from the Bootstrap-*t* (*B-t*) method seem to perform the best. Intervals from this method use a relatively full set of the information available from the bootstrap sample, but are neither constrained by distributional assumptions nor limited by the range of bootstrap results. In addition, these intervals have the advantage of being the most conservative. That is, intervals for this method allow for higher likelihood of impacts lower than the point estimate and give the lowest upper bounds. Thus, intervals for this method are least likely to overstate the local economic benefits of recreation visitation to the Florida Keys.

In this example, the lower bound for the 90-percent and 95-percent *B-t* confidence intervals are roughly 15 and 17 percent below the point estimate, respectively. The upper bounds for the same intervals are about 7 percent and 9.5 percent higher. These results were obtained from a sample of 505 expenditure

© Blackwell Publishers 2000.

observations. It is not known how different the intervals might be given a different sample size.

It is important to remember that the methods and results presented here do not represent the full range of variability in estimating the economic impacts of recreation. Rather, the confidence intervals are contingent upon the assumed form of the bridging and technical coefficients matrices. The true distribution of the elements in those matrices are unknown so standard practice is to assume values are fixed. The intervals are also contingent upon the assumption that the characteristics of the intercept sample are accurate and invariant. Still, the intervals presented here are conceptually comparable to intervals developed in research on recreation valuation. The same source of variation is incorporated in both economic measures. As well, similar methods are used to control variability from other sources.

This study does not incorporate any variation from estimates of the annual visitation to the Florida Keys. Confidence intervals were developed for impacts per 1,000 visitors, which are expected to be independent of the estimate of total visitation. Visitation estimates by access mode were available although standard errors for these estimates were not. If variance estimates for visitation were available, a confidence interval for total annual impacts may be developed following procedures for joint confidence intervals. Given a point estimate of 1.172 million visitors during the summer visitation to the Keys (Leeworthy 1996), total impacts would be about 1,195 million dollars (in 1993 dollars). Based on the *B-t* impacts results, a 90-percent confidence interval would cover the range of about 1,021 million dollars to 1,284 million dollars.

Further research is needed with regard to both empirical results and exploration of methods in generating approximate confidence intervals for economic impact measures of recreation. Either through compilation of a series of empirical bootstrap confidence interval results or through simulation studies, research is needed to examine relationships between characteristics of the original expenditure sample (for example, size, mean spending amount, length of stay) and width of confidence intervals. Collecting visitor expenditure data is often costly and understanding the tradeoffs between sample size and reliability of subsequent estimates may help researchers and managers better allocate their resources.

Adding approximate confidence intervals to impact analyses of recreation and tourism visitation is a worthwhile addition to standard analysis and reporting practices. Showing a likely range of impact estimates provides a great deal more information to planners and policy makers than is typically available. Comparability of information across both impacts and valuation provides benefits in itself, particularly for agencies whose objectives include both rural economic development and national benefits. Compared to the cost of conducting many economic impact studies, the additional cost of generating confidence intervals is justified in ensuring that the best use is made of the country's recreation and amenity resources.

© Blackwell Publishers 2000.

REFERENCES

Adamowicz, Wiktor L., Jerald J. Fletcher, and Theodore Graham-Tomasi. 1989. "Functional Form and the Statistical Properties of Welfare Measures," *American Journal of Agricultural Economics,* 71, 414–421.

Bergstrom, John C., R. Jeff Teasley, H. Ken Cordell, Raymond A. Souter, and Donald B. K. English. 1996. "Effects of Reservoir Aquatic Plant Management on Recreational Expenditures and Regional Economic Activity," *Journal of Agricultural and Applied Economics*, 28, 409–422.

Beyer, William H., ed., 1966. *Handbook of Tables for Probability and Statistics*. Cleveland: Chemical Rubber Company.

Blalock, Hubert M., Jr. 1979. *Social Statistics*, Revised 2nd Ed. New York: McGraw-Hill.

Cooper, Joseph C. 1994. "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys," *Land Economics,* 70, 111–122.

Creel, Michael D. and John B. Loomis. 1991. "Confidence Intervals for Welfare Measures with Application to a Problem of Truncated Counts," *The Review of Economics and Statistics,* 73, 370–373.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley & Sons.

Douglas, Aaron J. and David A. Harpman. 1995. "Estimating Recreation Employment Effects with IMPLAN for the Glen Canyon Dam Region," *Journal of Environmental Management,* 44, 233–247.

Efron, Bradley and Robert Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Johnson, Rebecca L. and Eric Moore. 1993. "Tourism Impact Estimation," *Annals of Tourism Research,* 20, 279–288.

Johnson, Richard A. and Dean W. Wichern. 1992. *Applied Multivariate Statistical Analysis*, 3rd Ed. Englewood Cliffs: Prentice Hall.

Kling, Catherine L. and Richard J. Sexton. 1990. "Bootstrapping in Applied Welfare Analysis," *American Journal of Agricultural Economics,* 72, 406–418.

Leeworthy, Vernon R. 1996. Technical Appendix: Sampling Methodologies and Estimation Methods Applied to the Florida Keys/ Key West Visitors Survey. Strategic Environmental Assessments Division, National Ocean Service, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Leeworthy, Vernon R. and Peter A Wiley. 1996. *Visitor Profiles: Florida Keys / Key West*. Strategic Environmental Assessments Division, National Ocean Service, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Leeworthy, Vernon R., Peter A Wiley, Donald B. K. English, and Warren Kriesel. 2000. "Correcting for Nonresponse Bias in Tourism Expenditure Surveys: An Application in the Florida Keys," *Annals of Tourism Research,* 28.

Miller, Ronald E. and Peter D. Blair. 1985. *Input-Output Analysis: Foundations and Extensions*. Englewood Cliffs: Prentice-Hall.

Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes. 1974. *Introduction to the Theory of Statistics*, 3rd Ed. New York: McGraw-Hill.

National Oceanic and Atmospheric Administration. 1995. Florida Keys National Marine Sanctuary, Draft Management Plan/ Environmental Impact Statement, Volume II. Washington, DC: U.S. Department of Commerce.

Park, Timothy, John B. Loomis, and Michael Creel. 1991. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies," *Land Economics,* 67, 64–73.

Propst, Dennis B., Daniel J. Stynes, Ju H. Lee, and Scott R. Jackson. 1992. Development of Spending Profiles for Recreation Visitors to Corps of Engineers Projects, Technical Report R-92-4, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Smith, V. Kerry. 1993. "Nonmarket Valuation of Environmental Resources," *Land Economics*, 69, 1–26.

© Blackwell Publishers 2000.

Souter, Raymond A. and James M. Bowker. 1996. "A Note on Nonlinearity Bias and Dichotomous Choice CVM: Implications for Aggregate Benefits Estimation," *Agricultural and Resource Economics Review*, 25, 54–59.

USDA Forest Service. 1990. A Strategic Plan for the 90's: Working Together for Rural America. Washington, DC: U.S. Department of Agriculture, Forest Service. June 1990.

USDA Forest Service. 1995. 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.

Uysal, Muzaffer, Robert S. Pomeroy, and Thomas D. Potts. 1992. "County-level Tourism Impact Assessment: A Case Study in Beaufort County, South Carolina," *Journal of Travel Research,* 31, 57–65.

Yen, Stephen T., and Wiktor L. Adamowicz. 1993. "Statistical Properties of Welfare Measures from Count-Data Models of Recreation Demand," *Review of Agricultural Economics*, 15, 203–215.

© Blackwell Publishers 2000.