Confidence Intervals for Welfare Estimators From Recreation Demand Models

Catherine L. Kling
and
Richard J. Sexton

February 27, 1989
University of California
Davis, California

"Paper to be presented at the Western Agricultural Economics Association meetings in Coeur d'Alene, Idaho, July 9-12,1989."

#6035

Confidence Intervals for Welfare Estimates From Recreation Demand Models

This paper analyzes the use of bootstrap methods to discern the statistical precision of consumer welfare estimates from recreation demand models. Results suggest that a great deal of imprecision is present in estimates from typical cross-section data. Precision can be improved through imposition of inequality constraints on the demand function parameters.

Confidence Intervals for Welfare Estimates From Recreation Demand Models

This paper investigates the use of bootstrapping procedures to estimate the statistical precision of point estimates of consumer welfare obtained from traditional recreation demand models. In a prototype recreation demand study, a functional form for demand is chosen and parameters of the model are estimated. The estimates are then used to calculate Marshallian consumer surplus or the more exact Hicksian measures of compensating or equivalent variation. Typically, no measure of the precision of the welfare estimator accompanies these point estimates. Finally, recommendations for policy or project decisions are based on the value of these point estimates.

Basing recommendations on welfare estimates about which confidence intervals are unknown does not seem to be a desirable state of affairs. This paper suggests improvements in this basic methodology using bootstrapping techniques. Bootstrapping is a technique for assessing the variability in an estimate using only the data at hand by "resampling" the original observations. A variation of the bootstrapping methodology has recently been introduced to applied welfare analysis by Graham-Tomasi, Adamowicz and Fletcher (G-TAF) and Adamowicz, Fletcher and Graham-Tomasi (AFG-T).

In this paper, we illustrate and analyze use of the bootstrap to construct a confidence interval around a point estimate of welfare. Unlike G-TAF and AFG-T, we analyze a variety of welfare estimators and use a Monte Carlo framework wherein we can evaluate the performance of the bootstrap in a variety of empirical settings. We also propose and evaluate an adaptation of the basic bootstrap methodology using inequality constraints that appears to address and reconcile some significant problems raised in the G-TAF and AFG-T analyses.

Consumer Welfare Estimators

To make the scope of this study manageable, we focus specifically upon the linear and semilog demand functions. These are the most commonly used functional forms in recreation analysis (McConnell and Bockstael and Strand (BS)), and both are compatible with utility maximization (Hanneman). The specific functions we work with are as follows:

(1) $X_i = a + bP_i + cY_i + u_i,$

(2) $X_i = \exp(a + bP_i + cY_i + u_i),$

where X_i is the quantity consumed of the relevant good by the i^{th} individual, P_i and Y_i are, respectively, his nonstochastic normalized price and income. The total Marshallian consumer surplus, $cs(P_i)$, measuring the entire area under the demand curve for a given price can be written $cs(P_i) = -X_i^2/(2\beta)$ and $cs(P_i) = -X_i/\beta$ for the linear and semilog demand functions respectively. Expressions for the consumer surplus associated with a small price change or deadweight loss can also be explicitly expressed as can the associated Hicksian measures (Bockstael, Hanemann and Strand (BHS)). In empirical work, the parameters and the disturbance vector must be replaced by statistical estimates to be denoted by $\hat{\theta}=\{\alpha, \beta, \gamma\}$, and $\epsilon = \{\epsilon_1,\ldots,\epsilon_n\}$, respectively.

Because α, β, and γ, are random variables, the various welfare estimators are also random variables and generally involve some form of ratio of random variables. As is well known (e.g., see Mood, Graybill, and Boes), no simple exact formulae exist for the mean and variance of ratios of random variables. Approximation formulae can be derived based on Taylor's series expansions. For example, for the estimated total Marshallian surplus, $c\hat{s}$, for the linear demand, a Taylor's series approximation to the variance can be written:

(3) $$\mathrm{Var}[c\hat{s}] \simeq \frac{X'^2}{b^2}\sigma^2 + \frac{X'^4}{4b^4}\mathrm{Var}\beta - \frac{X'^3}{b^3}\mathrm{Cov}(\beta,X),$$

where (3) is a first-order Taylor's series expansion about $E[X]=X'$ and $E[\beta]=b$.

It can be shown (Zacks) that (3) is the asymptotic variance of $c\hat{s}$. However, reliance on asymptotic formulae may generate significant errors in small samples (Freedman and Peters). Bootstrapping avoids the necessity of appealing to asymptotic formulae or making particular assumptions about the distributions of the estimated parameters.

Bootstrapping in Applied Welfare Analysis

Assessible introductions to bootstrapping are provided by Efron and Gong and Freedman and Peters. In the context of applied welfare analysis, the approach is as follows:

1. A functional form for consumer demand such as (1) or (2) is chosen. The disturbances u_i are assumed i.i.d. with mean zero. Normality of the u_i need not be assumed. The model is estimated on the sample data. The estimated parameters, α, β, and γ are used to compute the desired welfare estimate, $\hat{w}$.

2. The regression residuals ϵ_i are computed and used to create a distribution, ϕ. Each ϵ_i, i=1,...,n, has probability mass 1/n in ϕ.

3. New quantities $\{X_1^*,...,X_n^*\}$ are created from the formula $X_i^* = \alpha + \beta P_i + \gamma Y_i + \epsilon_i^*$, where ϵ_i^* is chosen by random draw with replacement from ϕ.

4. New parameter estimates, $\hat{\theta}^* = \{\alpha^*, \beta^*, \gamma^*\}$ are generated from regressing X^* on P and Y and used to compute a new welfare estimate, $\hat{w}^*$.

5. Steps 3 and 4 are repeated by redrawing from ϕ to generate an empirical distribution for $\hat{w}^*$. Its distribution characterizes the statistical uncertainty associated with $\hat{w}$.

Let η denote the distribution of the $\hat{w}^*$ generated from the above process. The Var$[\hat{w}]$ can be estimated by computing $E[\hat{w}^{*2}] = (1/T)\Sigma\hat{w}_t^{*2}$, $E[\hat{w}^*] = (1/T)\Sigma\hat{w}_t^*$, and invoking the usual formula:

$$\text{Var}[\hat{w}^*] = E[\hat{w}^{*2}] - (E[\hat{w}^*])^2,$$

where T is the number of bootstrap trials. The estimated variance should not be used to construct confidence intervals, though, because $\hat{w}$ is unlikely to be distributed normally in small samples. Rather, a $(1-\delta)$ percent central confidence interval around $\hat{w}$ can be constructed as simply as deleting the outer $\delta/2$ tails from η.

G-TAF modify the bootstrap procedure described above by assuming that the u_i are distributed normally and generating a vector of disturbances ϵ^* directly from a normal$(0,\hat{\sigma}^2)$ distribution rather than from the empirical distribution of the residuals, where $\hat{\sigma}^2$ is the sample variance. The merit of this modification is analyzed in our Monte Carlo analysis.

An apparent problem in the AFG-T and G-TAF pioneering analysis is that the bootstrap-generated data occasionally produce parameters yielding implausible welfare estimators. Specifically, draws generating absolutely small negative values for β may yield values of willingness to pay in excess of income and positive values for β may also be generated, implying negative willingness to pay. Such results are counter-intuitive and inconsistent with the economic theory underlying these models. Yet because these implausible estimators may be very large in absolute value relative to the magnitude of the other bootstrap-generated welfare estimates, they can easily cause bootstrap-estimated variances of welfare to appear to be very large.

The modification we propose and analyze here is to eliminate these aberrant bootstrap draws using inequality restrictions on the demand function parameters. Specifically, for any welfare measure, w, we require that $0 < \hat{w}_i < Y_i$, for all i=1,...,n, i.e., willingness to pay must be bounded by zero and total income. We also impose integrability conditions (BHS) on the demand functions which require $X_i < b/c$, for all i=1,...,n, for both the linear and semilog functions. These restrictions translate into inequality restrictions on the demand function parameter vector $\hat{\theta} = \{\alpha, \beta, \gamma\}$.

We impose the restrictions using Bayesian methods. The restrictions can be summarized in the prior density function for θ, $P(\theta)$. From Bayes Theorem, this prior density function, when combined with the sample likelihood function $L(\theta|X,P,Y)$, yields a posterior density function,

(4) $P(\theta|X,P,Y) \propto P(\theta)L(\theta|X,P,Y)$,

that is defined only over the parameter space consistent with the restrictions. Because the posterior density often cannot be integrated to obtain its moments, Monte Carlo integration (van Dijk and Kloek, Geweke) is used wherein a large number of draws of $\hat{\theta}$ are made from the posterior distribution and used to compute the moments of θ.

Our procedure differs from this approach only in that we used bootstrapped data sets, not random number generators to redraw $\hat{\theta}$. That is, each bootstrap generated parameter vector was checked to see if the inequality restrictions held. Those $\hat{\theta}^*$ that satisfied these restrictions comprised the empirical "posterior" distribution of $\hat{\theta}$ and were used, in turn, to create a posterior distribution of the w_i.

The Monte Carlo Experiment

The preceeding discussion has raised a number of problems and issues for applied welfare analysis including:

1. How is the precision of welfare estimators affected by the parameters of the data such as the sample size, the distribution and variance of the u_i, and the share of total consumer surplus relative to the total budget?

2. What is the relative size of confidence intervals around different welfare estimators and Hicksian vs. Marshallian measures?

3. Is the AFG-T simplification of the bootstrap methodology appropriate?

4. Is the tendency of the bootstrap to yield implausible welfare estimates ameliorated by introducing inequality restrictions?

The simulation experiment to analyze these issues involved construction of 16 data sets for both the linear and semilog functional forms. For each function, values were chosen for the parameters a,b, and c. Price (travel cost) and income data were taken from a Cheasapeake Bay beach survey conducted by Research Triangle Institute for the University of Maryland (BHS). Given values for a,b,c, and the P_i and Y_i, quantity variables, X_i, were constructed for each functional form using equations (1) and (2) by randomly drawing the u_i from an appropriate distribution.

The data sets are differentiated based on the following criteria: (1) distribution of the u_i--normal vs. uniform, (2) magnitude of the error variance--large vs. small, (3) sample size--n=25 vs. n=100, and (4) magnitude of the welfare measures--large vs. small, where large w_i were obtained by increasing the base values of a,b, and c. Each data set was characterized by a unique choice among these four criteria. Exhausting the possible combinations produced the 16 data sets for each functional form.

Point estimates α, β, and γ, and residuals ϵ_i were obtained for each data set using OLS. The estimated parameters and residuals were then used to produce point estimates of welfare; these estimates correspond to what an applied researcher would obtain. The final step in the simulation experiment was to conduct the bootstrap trials. Three alternative bootstrap procedures were employed on each data set: (1) the prototype bootstrap involving resampling from the empirical distribution of residuals, (2) the modified bootstrap with inequality restrictions, and (3) the AFG-T, normal-distribution bootstrap involving sampling the ϵ_i from a $N(0,\hat{\sigma}^2)$ distribution.

One hundred bootstrap trials were obtained for each procedure.

Results

Table 1 contains results from OLS estimation for the first eight of the 16 data sets described above. The final eight data sets involved relatively large welfare measures (criteria 4 above). For brevity, results for these data sets are omitted from the Tables. The Table also contains summary information on the characteristics of the data sets. The first four linear and semilog data sets are similar to typical recreation demand data; the R^2's are fairly low, ranging from 0.13-0.53, the t-statistics are generally significant, but not extremely large, and consumer surplus constitutes a relatively small part of the consumer's total budget.

To examine the relative performance of the three bootstrap methods in estimating the precision of point estimates, coefficients of variation (CV) were computed for each method for the total consumer surplus estimator (cs). CV is the standard deviation of the point estimate, $\hat{w}$, (as estimated by the applicable bootstrap method) divided by $\hat{w}$. Large CV imply imprecise point estimates and correspond roughly to large confidence intervals.

Table 2 contains the estimated CVs for cs using each of the alternative bootstrapping methods. The prototype bootstrap and the AFG-T normal bootstrap produce some extremely large CVs that are eliminated with the use of the modified bootstrap. For example, in the first linear model data set the omission of a single bootstrap trial results in a drop in the coefficient of variation from 27.98 to 0.97. Other cases of large improvements in the variance estimates also occur. These improvements are particularly likely to occur when the error variances are large or the number of observations is small. The improvement offered by the use of the modified bootstrap can be quantified by comparing the average CVs reported at the bottom of each Table. Use of the modified bootstrap induces a six-to-ten fold reduction in the reported variance of $\hat{cs}$.

To determine whether the smaller variance estimates in the modified bootstrap were due to the use of the emprical error distribution or the inequality restrictions, the experiment was repeated employing both normal errors and the inequality restrictions. Column 4 contains the results. With the exclusion of the aberrant bootstrap trials, sampling from a normal distribution generated similar results to those generated by sampling from the empirical distribution, an average CV of 0.55 from the normal vs. 0.40 from the empirical distribution for the linear model and 0.65 vs. 0.40 for the semilog. Even in cases where the true errors were uniformly distributed, sampling from a normal distribution did not appear to significantly alter the findings. These results suggest that researchers finding it more convenient to draw from a normal distribution can do so with confidence.

To examine the relative size of coefficients of variation across alternative welfare estimates, Table 3 reports CVs for the first eight data sets from the modified bootstrap for each of the six welfare estimators. These are total Marshallian and Hicksian consumer surplus (cs and CVT), small (5 dollar) price increase (CSS and CVS) and deadweight losses for the 5 dollar price increase (CSD and CVD). Several notable patterns emerge. First, despite the fact that the Hicksian measures are computed from estimates of Marshallian demand functions, rely on income coefficients as well as price coefficients, and have more complex formulae, the CVs are quite close between the Hicksian and Marshallian measures. Even the use of data sets 9-16 which produced noticeable discrepencies between the Hicksian and Marshallian measures did not result in significant differences in the CVs.

Second, the total consumer surplus and compensating variation measures are generally estimated less precisely than the other two measures, but the most significant difference appears to be in the estimation of the small price change consumer surplus. This result can probably be explained by noting the portion of each welfare estimate that relies on the precision of the regression estimates. For example, total consumer surplus is the area under the entire demand curve for a given price and, as a result, the measure is likely to be very susceptible to slight changes in the price coefficient. In contrast, the small price change consumer surplus measure contains a large rectangular portion which is non-stochastic. The deadweight loss measure is computed by subtracting the nonstochastic component from small consumer surplus, and, as such, it shares the total consumer surplus's reliance on the precision of the price coefficient estimate.

Finally, the CVs reported in Table 3 can be used to get a rough idea of the magnitude of the confidence intervals around the point estimates of welfare. If plus or minus two standard deviations is used as a rough approximation of 95% confidence intervals, then the average confidence interval around $\hat{cs}$ for both the linear and semilog models is approximately plus or minus 80%, with a range from 12 to 506%.

The size of these confidence intervals is disturbing, particularly for those data sets that are most similar to observed cross-sectional data. For example, the confidence intervals for the first four data sets average about plus or minus 220 and 185% of the cs for the linear and semilog models, respectively. However, in data sets 5-8, the confidence intervals are roughly plus or minus 27% and 34% for the linear and semilog models, respectively. The n=25 data sets also generate significantly larger confidence bounds (plus or minus 162 and 192% for the linear and semilog models compared to 84 and 26% for the n=100 data sets). There appear to be large returns to generating more observations and getting better model fits.

Also illuminating is to compare the standard error of $\hat{cs}$ to the Willig bounds on the difference between the Marshallian and Hicksian total welfare estimates. The ratio of two standard deviations of $\hat{cs}$ to the Willig bounds was computed for each of the sixteen point estimates of cs for the semilog model. The average of these ratios across the data sets is 774. The first 8 data sets where the difference between the Marshallian and Hicksian measures are very small yield, as expected, a much larger average ratio of about 1520. However, even when considering only the second 8 data sets where the differences between the Marshallian and Hicksian measures are much larger, the average ratio is about 14. These results suggest that there are far higher returns from improving the precision of our estimates than from worrying about calculating "exact" measures of welfare. One might say that there is a great deal of inexactness even in our exact measures.

Although the standard errors can be usefully used to generate crude confidence intervals, it is important to emphasize that intervals computed from the bootstrap distribution as described earlier will reflect any asymmetries in the small sample distribution of the $\hat{w}$ (Efron) and, hence, represent the preferred alternative. For example, for linear data set no. 1, the crude 90% confidence intervals is $\hat{cs} \in [0,951.35]$, while the estimated interval derived from the bootstrap distribution is $\hat{cs} \in [223.02,782.01]$.

Conclusions and Recommendations

This paper has analyzed the use of bootstrapping methods to estimate the statistical precision of common welfare estimators. Bootstrapping techniques can improve the state of the art of applied welfare analysis by enabling researchers to construct confidence intervals around welfare estimators without appealing to asymptotic properties or assumptions about underlying error distributions. Our simulations demonstrate that results are markedly improved through imposition of innocuous, theoretically-based inequality restrictions on the bootstrap-generated parameters. Imposing normality on the residuals did not, however, appear to be a significant restriction.

Our results also demonstrate that welfare measures tend to be estimated quite imprecisely, especially for data samples typical of applied recreation analysis. Standard errors for the total Hicksian and Marshallian measures in these cases were, on average, of the same magnitude as the point estimate, and the associated confidence intervals were much larger than the Willig bounds.

Finally, and perhaps, most instructively, the simulation results indicate that the standard errors of welfare estimators can be reduced, sometimes dramatically, with more data and better fits of the models. This result suggests that there may be greater returns to collecting more and better data than from worrying about theoretical discrepencies between various measures.

Table 1. OLS Estimation Results for the Linear and Semilog Data Sets

	<------- Description	-------->							
Data Set	σ_u^2	n	u	cs	β	t ratio	γ	t ratio	R^2
Linear Model									
1	lg	100	nor	sm	-0.07	-2.99	.00004	6.78	.35
2	lg	25	nor	sm	-0.06	-1.30	.00004	2.03	.15
3	lg	100	uni	sm	-012	-3.18	.00004	4.51	.22
4	lg	25	uni	sm	-0.15	-2.16	.00003	0.78	.13
5	sm	100	nor	sm	-0.09	-14.14	.00004	24.66	.89
6	sm	25	nor	sm	-0.09	-7.60	.00005	8.20	.85
7	sm	100	uni	sm	-0.09	-7.26	.00004	12.05	.66
8	sm	25	uni	sm	-0.10	-4.93	.00003	3.24	.60
Semilog Model									
1	lg	100	nor	sm	-0.03	-7.17	.000002	8.02	.53
2	lg	25	nor	sm	-0.02	-2.55	.000001	0.31	.17
3	lg	100	uni	sm	-0.03	-5.19	.000008	4.14	.30
4	lg	25	uni	sm	-0.01	-1.08	.000018	2.70	.23
5	sm	100	nor	sm	-0.03	-15.71	.000010	21.93	.88
6	sm	25	nor	sm	-0.02	-8.01	.000030	22.35	.96
7	sm	100	uni	sm	-0.02	-8.56	.000010	13.06	.71
8	sm	25	uni	sm	-0.02	-3.54	.000011	5.18	.63

Table 2. Coefficients of Variation for Total Consumer Surplus From the Alternative Bootstrap Methods

Data Set	Prototype	Modified		Normal	Normal Modified
Linear Model					
1	27.98	0.97	(101)	0.91	0.91
2	4.68	1.31	(118)	37.22	3.60
3	0.48	0.48	(106)	0.49	0.69
4	0.84	1.60	(211)	1.97	1.38
5	0.07	0.07	(100)	0.08	0.08
6	0.12	0.12	(100)	0.14	0.14
7	0.15	0.15	(100)	0.16	0.16
8	0.21	0.21	(100)	0.23	0.23
Mean[a]	2.49	0.41		3.87	0.55
Semilog Model					
1	0.16	0.16	(100)	0.15	0.15
2	0.81	0.79	(157)	4.36	4.40
3	0.18	0.18	(100)	0.21	0.21
4	2.46	2.53	(115)	32.11	2.66
5	0.06	0.06	(100)	0.06	0.06
6	0.12	0.12	(100)	0.14	0.14
7	0.12	0.12	(100)	0.12	0.12
8	0.39	0.39	(100)	0.45	0.45
Mean[a]	2.67	0.41		3.86	0.65

[a] Means are based on all 16 data sets.

Table 3. Coefficients of Variation for The Six Welfare Estimators

Data Set	CS	CSS	CSD	CVT	CVS	CVD
Linear Model						
1	0.97	0.03	0.37	0.99	0.03	0.38
2	1.31	0.06	0.61	1.34	0.06	0.62
3	0.48	0.04	0.27	0.48	0.04	0.27
4	1.60	0.08	0.42	1.61	0.08	0.42
5	0.07	0.01	0.07	0.07	0.01	0.07
6	0.12	0.01	0.11	0.12	0.01	0.11
7	0.15	0.01	0.14	0.15	0.01	0.14
8	0.21	0.02	0.20	0.21	0.02	0.20
Mean[a]	0.41	0.03	0.24	0.42	0.03	0.26
Semilog Model						
1	0.16	0.03	0.13	0.16	0.03	0.13
2	0.79	0.06	0.37	0.80	0.06	0.37
3	0.18	0.05	0.15	0.18	0.05	0.15
4	2.53	0.10	0.72	2.78	0.10	0.72
5	0.06	0.01	0.05	0.06	0.01	0.05
6	0.12	0.02	0.12	0.12	0.02	0.12
7	0.12	0.02	0.11	0.12	0.02	0.11
8	0.39	0.03	0.27	0.39	0.03	0.27
Mean[a]	0.40	0.07	0.23	0.45	0.07	0.24

[a] Means are based on all 16 data sets.

References

Adamowicz, W.L., J.J. Fletcher and T. Graham-Tomasi. "Functional Form and the Statistical Properties of Welfare Measures." American Journal of Agricultural Economics, 71(1989): in press.

Bockstael, N.E., W.M. Hanemann and I.E. Strand. "Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models", Final Report to the Environmental Protection Agency. EPA Contract No. CR-811043-01-0, 1986.

Bockstael, N.E. and I.E. Strand. "The Effect of Common Sources of Regression Error on Benefit Estimates." Land Economics, 63(1987):11-20.

Efron, B. "Nonparametric Standard Errors and Confidence Intervals." Canadian Journal of Statistics. 9(1981):139-158.

Efron, B. and G. Gong. "A Leisurely Look at the Boostrap, the Jackknife and Cross-Validation." American Statistician. 37(1983):36-48.

Freedman, D.A. and S.C. Peters." Bootstrapping a Regression Equation: Some Empirical Results." Journal of the American Statistical Association. 79(1984):97-106.

Geweke, J. "Exact Inference in the Inequality Constrained Normal Linear Regression Model." Journal of Applied Econometrics. 1(1986): 127-141.

Graham-Tomasi, T., W.L. Adamowicz, and J.J. Fletcher. "On Approximations to Expected Consumer Surplus." Staff Paper P88-44 Department of Agricultural and Applied Economics, University of Minnesota, Nov. 1988.

Hanemann, W.M. "Some Further Results on Exact Consumer Surplus." Working paper, Department of Agricultural and Resource Economics, University of California, Berkeley, 1982.

McConnell, K.E.."The Economics of Outdoor Recreation." Handbook of Natural Resource and Energy Economics, Allen Kneese and James Sweeney editors, Elsevier Science Publishers, 1985.

Mood, A., F. Graybill and D. Boes. 1974. Introduction to the Theory of Statistics. New York. McGraw-Hill Book Co.

van Dijk, H.K and T. Kloek. "Further Experience in Bayesian Analysis Using Monte Carlo Integration." Journal of Econometrics. 14(1980):307-328.

Willig, R.D. "Consumer's Surplus Without Apology." American Economic Review. 66(1976):589-97.

Zacks, S. The Theory of Statistical Inference. New York: John Wiley, 1971.