#9819

1991

Estimating the Precision of Welfare Measures

Catherine L. Kling

Department of Agricultural Economics
University of California, Davis
Davis, California 95616

AAEA, 1991

Welfare economics

February 12, 1991

Estimating the Precision of Welfare Measures

Three methods for constructing standard errors of welfare estimates have been employed in the recreation demand literature: a Taylor's series approximation, the bootstrap, and a method proposed by Krinsky and Robb. This paper presents the results of a simulation experiment designed to examine the accuracy of these methods.

Estimating the Precision of Welfare Measures

Researchers involved in applied welfare analysis have recently begun to stress the importance of providing estimates of the precision of welfare measures (Adamowicz, Fletcher, and Graham-Tomasi; Creel and Loomis; Kling and Sexton). A variety of approaches have been suggested, but since the standard errors of welfare estimates must themselves be estimated, the accuracy of these methods is of interest. The purpose of this paper is to consider alternative approaches to the estimation of standard errors and compare their performance under a variety of circumstances.

At least three approaches to estimating the standard error of a non-linear function of random variables have been proposed. The traditional approach linearizes the function via a Taylor's series approximation and computes an estimate of the standard error based on this linearized form (Kmenta). More recently, bootstrap methods have been proposed (Efron; Freedman and Peters) and applied to elasticities and consumer surplus (Dorfman, Kling, and Sexton; Green, Hahn, and Rocke; Kling and Sexton). In the bootstrap, the regression residuals are resampled to construct psuedo data sets which are used in turn to construct a distribution of the statistic of interest. A third method, proposed by Krinsky and Robb, is similar to the bootstrap, but the distribution of the estimated coefficients is resampled from directly.

To examine the accuracy of these three methods, this paper presents the results of a simulation study designed to examine the conditions under which each method is likely to provide accurate standard error estimates. A series of simulated data sets for which the parameters and error distribution are known are generated. This information is then used to construct the true standard error of Marshallian and Hicksian welfare estimates. Two sizes of welfare measures are examined: the value of loss of access to the good and the value of a change in quality of the good. Next, the steps a researcher might follow in doing a typical benefit estimation study are followed: the data sets are used to generate point estimates of welfare and each of the three methods is employed to estimate the standard error. The important difference between this study and a researcher faced with an actual data set is that the true standard error of the welfare estimates can be simulated; hence, this known standard error can be compared to the estimated standard errors to examine the accuracy of the method.

Problems and Methods in Estimating the Precision of Welfare Measures

To motivate the discussion, suppose that an individual consumer's true demand for recreation at a site is linear and can be written,

(1) $x_i = \alpha + \beta p_i + \gamma y_i + \delta q + \epsilon_i$,

where x_i is the number of trips individual i takes to the recreation site, p_i is the cost of accessing the site, y_i is the individual's income, q is the site quality, ϵ_i is an i.i.d. error term, and α, β, γ, and δ are non-random parameters.

The issue of precision in welfare estimation arises when the parameters of the demand function (1) are unknown; hence, must be estimated. Either Hicksian or Marshallian welfare measures can be computed based on the parameter estimates. Since the parameter estimates are random variables, so too, are the welfare estimates. The sum of the Marshallian and Hicksian surplus associated with the total access to the site over all individuals in a sample divided by the sample size (n) is the typically reported point estimate of average surplus, denoted cs and cv,

$$cs = (1/n)\Sigma_i cs_i = (1/n)\Sigma_i (1/2) x_i^2/\hat{\beta},$$

(2) and

$$cv = (1/n)\Sigma_i cv_i = (1/n)\Sigma_i[-\hat{\beta}/\hat{\gamma}^2 + \exp(-\hat{\gamma}x_i/\hat{\beta})(x_i/\hat{\gamma} + \hat{\beta}/\hat{\gamma}^2)],$$

where the hats indicate estimates of the corresponding parameters in (1).

A first order Taylor's series (TS) approximation to the statistics in (2) can be used to construct an estimate of their standard errors. The statistic of interest is first approximated via Taylor's series and the standard error of the linear approximation is constructed. This method is also referred to as the delta method and it produces asymptotic standard errors. Kmenta provides a good description of the procedure.

A second method for estimating the standard error is the bootstrap (BS) attributable to Efron. In the linear case, the estimated demand function is

(3) $x_i = \hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i + \hat{\delta}q + \hat{\epsilon}_i$, i=1,...,n.

To perform the bootstrap, the empirical distribution of the residuals, the $\hat{\epsilon}_i$'s, are randomly sampled with replacement and used to generate a new vector of x_i's, i.e.,

(4) $x_i^* = \hat{\alpha} + \hat{\beta}p_i + \hat{\gamma}y_i + \hat{\delta}q + \hat{\epsilon}_i^*$, i=1,...,n,

where $\hat{\epsilon}_i^*$ is a residual randomly drawn from the distribution of the $\hat{\epsilon}_i$ and x_i^* is the resulting bootstrap quantity vector.

These new quantities are used to re-estimate the demand function resulting in bootstrap estimates of the parameters, $\hat{\alpha}^*$, $\hat{\beta}^*$, $\hat{\gamma}^*$, and $\hat{\delta}^*$. The starred values are then used to compute the statistic of interest, namely cs or cv. This process of resampling from the empirical distribution of the errors, constructing psuedo data and re-estimating demand and welfare measures is repeated a large number of times. Each trial results in bootstrapped welfare measures, denoted cs^* and cv^*. The standard error of the starred welfare estimates is an estimate of the standard error of cs and cv.

The third method of estimating standard errors is similar to the bootstrap. However, this method, proposed by Krinsky and Robb (KR), resamples from the coefficient estimates directly. Denote the estimated coefficient vector, $\hat{\theta}$, and its estimated variance-covariance matrix, VC. Then, the procedure is to take random drawings from a multivariate normal distribution with a mean of $\hat{\theta}$ and variance-covariance matrix VC, generating new coefficient values $\hat{\theta}^*$. For each draw from this distribution, the welfare measures are computed. Again, denote the welfare measures computed from the $\hat{\theta}^*$ as cs^* and cv^*. A large number of draws results in a distribution of the cs^* and cv^* and the standard error of this distribution represents an estimate of the standard error of cs and cv.

One might expect that the bootstrap will provide more accurate estimates than the other methods when the underlying error distribution is non-normal, particularly when it is highly skewed. There may also be differences in the methods' accuracy in small samples or when the estimated demand function does not fit the data well, as measured by t-statistics or R^2. These issues will also be addressed in the simulation study.

Data Construction and the Simulation Study

To evaluate the performance of the three methods in estimating the standard error of welfare estimates, a simulation experiment is performed. First, a simulated data set is constructed by combining data on prices, income, and quality with parameter values and an assumed error distribution. This yields simulated quantities of trips. This information, prices, incomes, site qualities, and quantities constitutes a simulated data set. Point estimates of cs and cv are computed by estimating a demand function as if this were observed data.

The data on incomes and prices are taken from Chesapeake Bay beach users surveyed in the summer of 1984 by the Research Triangle Institute for the University of Maryland (see Bockstael, Hanemann, and Strand). Parameter values were chosen to provide site visitation rates that were similar to observed data. The values of the parameters are: $\alpha = 5.0$, $\beta = -0.01$, $\gamma = 0.00004$, and $\delta = 0.06$.

Since the parameters are known, as is the distribution of the error term, it is possible to simulate the standard error of welfare estimates, cs and cv. (That is, generate through simulation the "true" standard error of cs and cv). To do this, errors are redrawn from the assumed error distribution creating new quantities; these new quantities are then used to estimate the model yielding new point estimates of welfare. This is conceptually equivalent to drawing a new sample from the population. An empirical distribution of the estimated welfare measures can be constructed by repeating the procedure a large number of times and the standard error of the simulated welfare measures can be computed from this distribution.

Note the similarity of this procedure to the bootstrap described above. The key difference is that the true error distribution and true parameter values are used for resampling, whereas the bootstrap must resample from the empirical distribution and parameter estimates since the truth is unknown. Thus, the actual distribution of the welfare estimates is formed in the first case, but only an estimate in the second.

Having constructed a simulated data set, point estimates of welfare and the standard error of the welfare estimates, the next step in the simulation methodology is to apply each of the three standard error estimation methods to the data. Since the performance of the three standard error estimators may differ under different conditions, a series of simulated data sets is constructed and the simulation experiment outlined above performed on each one. The various data sets are differentiated based on sample size, distribution of the error term, ϵ_i, and the range of the error term (i.e., the variance of ϵ_i).

Three sample sizes (200, 100, and 50) and three error distributions, (normal, uniform, and χ^2) are employed in constructing the data. In addition, two relative sizes of the standard error of the error distributions are employed; in the first, the error distribution range is about (-1.5, 1.5), and in the second, the range is about (-3.0, 3.0).[1] In total, 18 simulations based on a linear demand function are performed, exhausting all of the combinations of these three characteristics.

For each of the 18 simulations, standard error estimates of Marshallian consumer surplus and Hicksian compensating variation associated with access to the site and associated with a doubling of site quality were constructed using the TS, KR and BS procedures. To compute the standard error estimates using BS and KR, 100 repetitions were performed.

In each simulation, 100 realizations of the simulated model are examined; that is, 100 drawings of the ϵ vector were made, creating 100 simulated data sets, and 100 TS, KR, and BS estimates of the standard error of the welfare measures. In total, 1800 data sets were created, 18 simulations x 100 realizations per simulation.

An Evaluation of the Methods

To examine the circumstances under which each method performs well, the absolute value of the percentage errors (PE)[2] between the estimated standard error

---

1. The standard error for the normal distribution small case is 0.25, the large is 1.0. For the uniform, the errors range from -1.5 to 1.5 for the small and from -3.0 to 3.0 for the large. The χ^2 has 2 degress of freedom and is multiplied by a constant to generate a standard error of 0.25 in the small case, and 1.0 in the large, and is shifted to have mean zero.

2. The percentage error is computed for TS simply as $(\hat{\sigma}_{TS} - \sigma)*100/\sigma$, where σ is the true standard error and $\hat{\sigma}_{TS}$ is the standard error as estimated by the TS. An equivalent expression is used to compute the PE for KR and BS by substituting in the appropriate standard error estimate.

and the simulated standard error is computed and reported in tables 1 and 2 for the total and quality change consumer surplus, respectively. The numbers reported are averages over the 100 realizations of each simulation. The first three columns in the table report the characteristics of the data: the distribution of the error term in the demand function, the sample size, and the relative range of the error distribution (a large or small error standard error). The fourth through sixth columns contain the absolute value of the percentage error. The table contains results for only the Marshallian measures since the results for the Hicksian measures are nearly identical.

The bottom of tables 1 and 2 contain overall means of the PE's of the standard error estimates, as well as means broken down by characteristic of the data sets. On average, TS has the lowest PE for the total cs and KR has the smallest for the quality change cs. In both cases, the KR and BS are extremely close: 19.76 vs. 19.60 in the total cs case and 20.66 vs. 20.81 in the quality change case.

To examine the effect that smaller sample size has on the accuracy of the standard error estimates, the mean of the percentage errors are computed separately for each sample size and reported below the overall means in the tables. For sample sizes of 200, all three methods estimate the standard error of the total cs quite accurately, falling within 6 percentage points of the truth. The quality change cs is apparently more difficult for the methods to estimate; the error ranges from 14% (for BS) to 27% (for TS). The percentage error jumps sharply when the sample size falls from 200 to 100 for all three methods for the total cs case, but surprisingly the PE doesn't increase much for TS and KR in the quality change case, and actually falls for BS. However, in moving from 100 to 50 observations, all of the methods for both size welfare changes yield a large loss in accuracy as measured by the PE.

The effect of the error distribution on the PE is captured in the next set of means in each table. In general, the PE is much larger for the uniform error case than for the normal or χ^2 cases. It appears that on average the methods have more difficulty with wide tailed distributions (the uniform) than either bell shaped (the normal) or skewed to one side distributions (the χ^2). Quite surprisingly, these results suggest that the bootstrap is not better than the KR procedure at estimating the standard error of distributions resulting from residuals with highly non-normal errors. In fact, the KR and BS generate very similar PEs for both the total and quality change cases. Since the non-distributional assumption in the bootstrap is one of its main advantages, it may be that the simpler TS or KR procedures can often be used in its place. However, the TS does perform noticeably better when the underlying error distribution is normal.

The effect of the error variance is examined in the final set of means. In the small error case, the average regression R^2 is about 0.81, and in the large residual case the average is about 0.55. When the range of the error doubles (moving from the small to large case), the PE more than quadruples for all three methods in the total cs case. The increase is just as dramatic in the quality change case for KR and BS where the PE increases by over seven times. However, for the TS the increase is much smaller since the TS performs poorly even in the small error case. These results dramatically point out that the goodness of fit affects the accuracy of the methods.

Finally, an interesting question to examine is whether the methods have a tendency to over or underestimate the standard error. In particular, the Taylor's series has been accused of this tendency by several authors (Green, Hahn, and Rocke; Krinsky and Robb). To examine this, the difference between the standard error estimates and the true standard error were examined. The tendency on the part of TS to underestimate is confirmed in these simulations. In a total of 31 of the 36 combined total cases, TS underestimates the standard error; this understatement was particularly pronounced in the quality change case where all of the standard error estimates were too small. This suggests that the tendency of the Taylor's series to underestimate may depend on the statistic of interest. Both BS and KR indicated a tendency to overstate the standard error (KR overstated in 30 of the 36 cases and BS overstated in 32).

Final Remarks and Conclusions

The precision with which welfare measures are estimated is of great importance to policy makers and researchers alike. This study has undertaken to provide an assessment of three methods of estimating the precision of welfare measures: the Taylor's series, the bootstrap, and a method suggested by Krinsky and Robb.

The bootstrap and the Krinsky and Robb procedure in most cases produce quite similar standard error estimates. This was true even when the underlying error distribution was non-normal. This similarity in the methods suggests that the less expensive Krinsky and Robb procedure may often be fruitfully substituted for the full fledged bootstrap.

All of the methods are considerably more accurate with sample sizes near 200 than with the smaller sizes of 100 and particularly 50. This is not surprising, but does reinforce the need for adequate sample sizes when performing welfare analysis. The TS performs best when the underlying error distribution is normal; the KR and BS were less sensitive on average to the error distribution, although both performed worse with the uniform distribution.

The TS underestimated the standard error quite often; in contrast, the KR and BS tended to overstate the standard error. Since overstating the standard error would seem to more desirable than understating it, researchers may wish to choose the KR or BS methods on these grounds. The simulation also indicated that there is little difference in the estimated standard errors of Marshallian consumer surplus and Hicksian compensating variation of welfare measures using either the TS, KR, or the BS. In most cases, all three methods provided reasonable approximations to the standard errors.

An important qualifier to the results presented here is appropriate. In applied recreation demand studies, researchers typically employ more complex estimators or functional forms than those used here. Applications using censored or truncated models, discrete choice methods, count data, flexible functional forms, or other complications may not experience the same accuracy of the standard error estimation methods as found in this set of experiments. For these more complex models, the choice of technique may be dominated by analytic or computational advantages of one method over another.

Table 1

Mean Absolute Percentage Errors of Marshallian Total Consumer Surplus Standard Deviation Estimates

Distribution	Sample Size	Error Variance	Taylor's Series	Krinsky & Robb	Bootstrap
normal	200	small	3.97	2.89	4.18
normal	100	small	3.10	6.57	6.31
normal	50	small	0.52	5.97	5.29
uniform	200	small	0.09	4.47	5.73
uniform	100	small	6.36	4.77	4.38
uniform	50	small	4.20	24.93	32.64
χ^2	200	small	1.67	2.60	3.64
χ^2	100	small	1.20	2.96	4.28
χ^2	50	small	4.51	1.31	3.08
normal	200	large	1.18	7.96	7.57
normal	100	large	2.71	17.87	11.75
normal	50	large	10.65	52.83	61.86
uniform	200	large	17.57	16.31	6.45
uniform	100	large	36.96	70.47	46.64
uniform	50	large	49.79	47.17	63.41
χ^2	200	large	0.87	6.73	6.80
χ^2	100	large	0.65	27.69	27.43
χ^2	50	large	72.62	52.21	51.51
Mean			12.15	19.76	19.61
Mean	n = 200		4.22	6.83	5.73
	n = 100		8.50	21.72	16.80
	n = 50		23.71	30.74	36.30
Mean	normal distribution		3.69	15.68	16.16
	uniform distribution		19.16	28.02	26.54
	χ^2 distribution		13.59	15.58	16.12
Mean	small error variance		2.85	6.27	7.72
	large error variance		21.44	33.25	31.49

Table 2

Mean Absolute Percentage Errors of Marshallian Quality Change Consumer Surplus Standard Deviation Estimates

Distribution	Sample Size	Error Variance	Taylor's Series	Krinsky & Robb	Bootstrap
normal	200	small	28.74	1.79	3.47
normal	100	small	24.62	0.14	0.65
normal	50	small	23.57	0.71	1.48
uniform	200	small	28.00	2.55	3.61
uniform	100	small	26.89	3.85	3.85
uniform	50	small	21.50	21.43	26.02
χ^2	200	small	23.23	5.48	5.68
χ^2	100	small	21.92	2.71	4.19
χ^2	50	small	23.71	1.52	1.74
normal	200	large	22.72	58.38	57.80
normal	100	large	24.08	14.19	9.21
normal	50	large	27.12	31.54	47.72
uniform	200	large	34.21	18.45	7.25
uniform	100	large	46.50	60.65	37.29
uniform	50	large	46.97	61.19	76.68
χ^2	200	large	23.99	7.46	8.01
χ^2	100	large	20.62	23.31	23.28
χ^2	50	large	77.59	56.49	56.65
Mean			30.33	20.66	20.81
Mean	n = 200		26.82	15.69	14.30
	n = 100		27.44	17.47	13.08
	n = 50		36.74	28.81	35.05
Mean	normal distribution		25.14	17.79	20.06
	uniform distribution		34.01	28.02	25.78
	χ^2 distribution		31.84	16.16	16.59
Mean	small error variance		24.69	4.47	5.63
	large error variance		35.98	36.85	35.99

References

Adamowicz, W., J. Fletcher, and T. Graham-Tomasi, Functional Form and the Statistical Properties of Welfare Measures, American Journal of Agricultural Economics 71, 414-421 (1990).

Bockstael, N., W. Hanemann, and I. Strand, "Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models: Volume II of Benefit Analysis Using Indirect or Imputed Market Methods", EPA Cooperative Agreement CR-811043-01-0, (1986).

Creel, M. and J. Loomis, Confidence Intervals for Marshallian and Hicksian Welfare Measures with Application to a Problem of Truncated Counts, manuscript, University of California, Davis (1990).

Dorfman, J., C. Kling, and R. Sexton, Confidence Intervals for Elasticities and Flexibilities: Re-Evaluating the Ratio of Normals Case, American Journal of Agricultural Economics. 72,1006-1017 (1990).

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

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

Green, R., W.Hahn, and D. Rocke, Standard Errors for Elasticities: A Comparison of Bootstrap and Asymptotic Standard Errors, Journal of Business and Economic Statistics, 5, 145-49 (1987).

Kling, C. and R. Sexton, Bootstrapping in Applied Welfare Analysis, American Journal of Agricultural Economics, 72, 406-418 (1990).

Kmenta, J. "Elements of Econometrics", MacMillan Publishing Co., Inc., New York, (1971).

Krinsky, I. and A. Robb, On Approximating the Statistical Properties of Elasticities, Review of Economics and Statistics, 68, 715-719 (1986).