April 1992 No. 338

A Convolutions Approach to Measuring the Differences in Benefit Estimates from Dichotomous Choice Contingent Valuation Studies

Gregory L. Poe*

Eric K. Lossin*

Michael P. Welsh**

* Research Assistants - Department of Agricultural Economics, University of Wisconsin - Madison. ** Senior Project Manager, HBRS.
The comments and support of Richard C. Bishop are gratefully acknowledged. Funding for this paper was provided by Hatch grant 142-d297. Data used in this paper was collected for Glen Canyon Environmental Studies sponsored by the U.S. Bureau of Reclamation.

A Convolutions Approach to Measuring the Differences in Benefit Estimates from Dichotomous Choice Contingent Valuation Studies

Dichotomous choice formats are now applied frequently in contingent valuation studies, but previously developed significance tests for differences in empirical distributions of benefit measures have either invoked normality assumptions or used non-overlapping confidence interval criteria. This paper demonstrates that such methods will generally not be appropriate, and develops an exact empirical test, based on the method of convolutions, for assessing the statistical significance between distributions of dichotomous choice contingent valuation welfare estimates. Application of the proposed convolutions approach is illustrated in a case study using two alternative techniques of generating empirical distributions from dichotomous choice data.

A Convolutions Approach to Measuring the Differences in Benefit Estimates from Dichotomous Choice Contingent Valuation Studies

Introduction

The dichotomous choice contingent valuation method (DC-CVM) has a number of widely discussed advantages over other contingent valuation elicitation techniques such as payment cards, bidding games, and open ended/direct questions [31, 9]. Most notably, the take-it-or-leave-it format of DC-CVM closely resembles decisions regularly made in the marketplace or voting booth, is easily incorporated into mail surveys, and is thought to impose less burden on respondents. This approach is incentive-compatible and may be less subject to strategic response than other elicitation methods [20, 29].

One drawback of the dichotomous choice technique is its relative inefficiency in collecting information about individual values [18, 12]. A "Yes" or "No" to a dichotomous choice question merely provides an indication of the direction of the relationship between the hypothetical cost or 'bid' value and the individual's true willingness to pay. A "Yes" response, for example, indicates that the suggested cost is less than the actual value that the individual places on the project.

Since willingness to pay is not directly elicited from each individual more complex statistical techniques are required to infer the average or total Hicksian surplus value.[1] Data generated by the dichotomous choice technique are generally analyzed using qualitative response models based on normal or logistic density functions. Estimates of the surplus value are recovered from the estimated parameters in the qualitative response model or by numerical integration. The precision of these welfare estimates has, until recently, been largely unknown, a fact which hinders direct statistical comparisons of benefit estimates obtained using different populations or hypothetical scenarios. Such comparisons are essential to assessing the validity and reliability of the dichotomous choice method. The ability to statistically compare distributions would also be important to assessing the impact of information, embedding, intertemporal considerations and other factors on contingent values. Previously, analyses of these impacts have been relegated to other valuation techniques (e.g., 2, 8, 23, 28).

As a proxy for variance estimates, early studies into the application and the validity of the dichotomous choice technique (e.g., 3, 4) relied on traditional statistical tests of the estimated coefficients to provide an indicator of the robustness of their derived welfare measures [13]. Log likelihood ratio tests were used to test the hypothesis that the estimated coefficients were equal across elicitation methods and estimation techniques. This approach has some very obvious limitations. First, even with a single population or elicitation technique, goodness of fit statistics do not necessarily translate into precision of the welfare estimates[2]. For example, a very 'flat' response function across bids may satisfy goodness of fit statistics but provide a willingness to pay estimate with a high coefficient of variation. Statistical differences across regression equations are also not sufficient to conclude that the equations will provide different estimates of willingness to pay. It is possible that two significantly different response functions, that are not stochastically dominant in the first degree, will provide the exact same estimate of willingness to pay.

Taylor series or Delta Method approximations of variance circumvent some of the limitations associated with the above approach. However, it is questionable whether these approximations will appropriately characterize the distributions of many of the benefit measures which typically involve complex non-linear functions of parameters. Indeed, evidence from other studies suggest that the linear approximations of variance can provide biased estimates of the variance of functions of estimated parameters [25, 26] In addition, linear approximations can not be applied to estimators of mean willingness to pay that do not have closed-form solutions (e.g., 10).

Recently, Cameron [13] offered an alternative to these techniques by providing an analytical solution for the variance of median willingness to pay from logit and probit models (see also Kealy, Montgomery and Dovidio [24]). A caveat to this analytical approach is that the median, which equals the non-truncated mean for symmetrical distributions, is not universally accepted as a welfare measure among contingent valuation practitioners. Quite simply, the use of the median as a welfare measure is not consistent with Pareto efficiency [22, 18, 19, 10, 15]. Moreover, it is not uncommon for median estimates to have negative values, a result that is somewhat "disturbing" and "unreasonable" for most resources [7].

Empirical 'simulation' techniques offer another method to estimating confidence intervals for welfare measures [27, 33, 25, 1, 15]. In contrast to the Cameron approach these techniques are more universal in that they can be applied to any function of the parameters estimated in logit or probit models. Furthermore, empirical approaches are able to capture the non-linear nature of most welfare estimates and are equally applicable to welfare measures that do not have a closed-form solution. While these methods are very computer intensive, they are becoming increasingly popular among contingent valuation practitioners.

Although these empirical techniques have provided great insights into the distribution and precision of point estimates of welfare measures derived from DC-CVM, relatively little effort has been devoted to applying these methods to exploring the difference between distributions of welfare measures based on different samples or estimation techniques. The objective of this paper is to develop a statistical approach, based on the method of convolutions, that provides an exact estimate of the significance of the difference between two empirical distributions. This technique relies only on the empirical distributions already calculated in the estimation of confidence intervals of individual point estimates and therefore offers a ready extension of existing and future techniques of generating empirical distributions.

The remainder of the paper is organized as follows. To provide a beginning point for the analysis the second section briefly reviews and discusses empirical methods that have been previously applied to estimating confidence intervals for DC-CVM. The third section provides a critique of methods that have either been implicitly or explicitly suggested for comparing the empirical distributions of point estimates. The proposed convolutions approach to comparing empirical distributions of welfare estimates is developed in the fourth section. This technique is applied in the fifth section to the evaluation of differences in willingness to pay for alternative water flow scenarios for Grand Canyon white-water boaters.

Methods for Approximating Empirical Distributions

Duffield and Patterson (DP) and Park, Loomis and Creel (PLC) have recently suggested two alternative techniques of estimating empirical distributions for benefit estimates from DC-CVM. While these techniques share the same philosophy of resampling, the two methods differ quite a bit in application. The DP approach employs a monte-carlo or bootstrapping technique that samples from the estimated binomial distributions for each bid level to create a 'pseudo data set' for which new parameter coefficients are estimated. This process is repeated a large number of times to generate an approximate empirical distribution of the desired welfare measures. PLC apply Krinsky and Robb's (KR) method that instead samples from the variance-covariance matrix to create approximate empirical distributions of welfare measures calculated from logit parameters.

Some underlying concepts and notation need to be provided in order to apply these models and to evaluate their differences. The dichotomous choice model presented here is motivated by a tolerance or expenditure difference approach which provides a basis for the DP analysis [12, 18, 19, 15]. Alternatively, as is presented in PLC, an indirect utility difference approach can also be invoked [18, 30]. To simplify this presentation the bid value is assumed throughout to be the only independent variable.

An assumption underlying the tolerance approach is that each individual has a true maximum value that he or she places on a resource or proposed project. Let the distribution of this value for the population, which shall be represented by willingness to pay (WTP), be characterized by the cumulative density function G(A) where A is a continuum of dollar amounts. Within this framework, the probability of a "Yes", π(A), response to a bid value, A, is given by

$$\pi(A) = Pr(true\ WTP > A) = 1 - G(A) \tag{1}$$

Estimation of π(A) (or G(A)) is accomplished by distributing bids $A_1, \ldots A_m$ across $1, \ldots, n$ individuals with $m \leq n$. For each individual a Yes/No response (r_n) is obtained. This sampling strategy provides a data set of n observations (n, A_m, r_n), from which π(A) is estimated non-parametrically (as in Kriström [27]) or parametrically by assuming some underlying distribution. Most commonly in DC-CVM the following standard logistic distribution is assumed

$$\pi(A) = [1 + e^{-(\alpha + \beta A)}]^{-1} \tag{2}$$

which produces estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\pi}$(A) and an estimated variance covariance matrix $\hat{\Sigma}$. For notational convenience, let $\hat{\underline{\beta}} = [\ \hat{\alpha}\ \hat{\beta}]$.

In turn, $\hat{\pi}$(A) can be used to estimate desired welfare parameters. For example, mean willingness to pay in DC-CVM has been shown to be:

$$E(WTP) = \int_0^{\infty} \pi(A)\, dA \tag{3}$$

where E is the expectation operator [19].

Within this framework, the DP approach to estimating an approximate empirical distribution involves three steps.

1. Using a random number generator, a pseudo-response vector ($\underline{r}^* = r_1^*,\ldots,r_n^*$) is generated by letting $r_1,\ldots,r_n$ be observations from independent bernoulli distributions for each $\hat{\pi}(A_n)$.
2. Reestimate equation (2) with $\underline{r}^*$ to obtain $\underline{\beta}^*$ and $\pi^*(A)$, and other desired values such as E(WTP) from equation (3).
3. Repeat steps 1 and 2 a large number (B) of times to obtain an empirical distribution of the desired parameters.

As such, the DP method is a monte-carlo approach.

The KR simulation method is a resampling approach that uses the information on the distribution of the estimated coefficients provided in the variance-covariance matrix (Σ) to create an empirical distribution and confidence interval for a function of the estimated parameters. The essence of the approach, as applied to DC-CVM, is as follows. Assuming that the true parameter vector $\underline{\beta}$ has a multinomial normal distribution, $N(\underline{\beta},\Sigma)$, a series of random draws (B') is made from the approximated distribution $N(\hat{\underline{\beta}},\hat{\Sigma})$. For each drawing, ($\hat{\hat{\beta}}_i$), a new $\hat{\pi}(A)$ is estimated, along with other desired values such as E(WTP) from equation (3). Empirical distributions of the desired estimates are generated using the complete set of random draws. PLC argue that this technique is relevant to DC-CVM because variance-covariance matrices of the maximum likelihood estimation methods typically used in estimating the coefficients in equation (2) have asymptotic multinomial normal properties[3].

If we accept the estimates of β and Σ and the multinomial normal distributional assumptions then an infinite number of draws should lead to the precise distribution of the desired function of the estimated parameters using the PLC technique. The same result is predicted for the DP method. In practice, however, a trade-off between accurate estimation of the empirical distribution and computational time is recognized, and B and B′ are determined by the number of repetitions that provide stable endpoints for the confidence intervals. Both PLC and DP suggest that 1000 repetitions will provide a reasonably accurate measure of confidence intervals.

Once the distributions have been estimated, there are two possible avenues for approximating a $(1-\alpha)$ confidence interval. The first approach is to assume that the empirical distribution is normal and construct a standard two-sided confidence interval. As will be discussed later, this normality assumption maybe somewhat naive in many cases. A second 'percentile' based approach is to construct an approximate $(1-\alpha)$ confidence interval by dropping $(B*\alpha/2$ or $B'*\alpha/2)$ observations from each tail.

At this point in time, it is not clear which of these techniques is superior. Both have certain advantages and disadvantages. The advantage of the PLC approach is that coefficients are sampled directly, and, thus, it requires relatively little computing power to construct approximate empirical distributions. This contrasts with the DP method which requires reestimation of the coefficients for each replication, an important fact when iterative estimation processes do not converge rapidly. On the other hand, the PLC assumption that the variance-covariance matrix is asymptotically multinomial normal distributed may not be relevant to estimates from small samples.

A Critique of Past Methods for Evaluating Differences in DC-CVM Welfare Estimates

Given approximate distributions and confidence intervals for welfare estimates, the question arises about how to statistically compare these approximate distributions. Two techniques for comparing empirical distributions have been proposed in the literature. The first, as implicitly suggested by Krinsky and Robb, is that if the simulated distributions are approximately normal then classical statistical procedures for estimating differences can be applied. For instance, assuming equal and known standard errors (STDERR) for two normal distributions, the null hypothesis that the 'true' mean of the first distribution (μ_1) is equal to the 'true' mean of the second distribution (μ_2) is tested using the following difference formula,

$$Z=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{2}*STDERR} \qquad \sim N(0,1) \tag{4}$$

where z is the test statistic, $\overline{X}_i$ are the sample means, and n is the number of observations [36, p. 101]. As noted, the Z value has a standard normal distribution. In the framework of PLC and DP, the standard error of the mean is approximated by the estimated standard deviation of the empirical distribution of the mean estimate [15].

Based on our experience with estimating mean welfare measures from logistic functions, empirical distributions are skewed, and these 'approximately normal' conditions do not hold. This observed asymmetry is especially true for estimators of mean willingness to pay such as equation (3) in which the range of integration is truncated at a lower bound of zero. More generally, there is little reason to assume that non-linear functions of normal parameters will approximate a normal distribution.

PLC avoid this issue by employing a non-overlapping confidence interval criterion to evaluate differences in point estimates. That is, the differences in mean willingness to pay across estimations are judged to be statistically significant at the 5 percent level if their empirical 95 percent confidence intervals do not overlap. In general, the actual significance of this non-overlapping confidence interval approach will not correspond to the stated level of the test. This point is demonstrated most simply for normal distributions using the analytical solution presented in equation (4). Recall that for a single normal distribution the 95 percent confidence interval for the mean of an estimate is defined as $\overline{X}_i \pm 1.9600*(STDERR)$. Again assuming that the standard errors for both distributions are known and equal, this implies that the critical difference in means, $(\overline{X}_1 - \overline{X}_2)$, associated with the non-overlapping 95 confidence intervals would have to be at least 3.9200 standard errors apart before they would be judged to be significantly different. Making this substitution, equation (4) becomes

$$Z=\frac{(3.9200*STDERR)}{\sqrt{2}*STDERR}=\frac{3.9200}{\sqrt{2}}=2.772 \qquad \textbf{(5)}$$

The estimated z value of 2.772 corresponds to a significance level (which shall be referred to as α') of 0.0048 rather than the stated value of α=0.05[4].

Clearly the non-overlapping confidence interval criterion given by $(1-\alpha)$ confidence intervals does not correspond to the α level of significance for the normal case. In general, a lack of correspondence between α and α' is expected. For the normal distribution above, the significance level is understated (i.e. $\alpha>\alpha'$) and the test is more conservative than indicated. The degree of this difference between α and α' will depend upon the shape of the empirical distributions that are being compared.

The Method of Convolutions

Another alternative - one that accommodates any distributional form - is based on the method of convolutions. This method is a well established technique used in statistics and mathematics to evaluate the sum of distributions of random variables and series [17, 32].

Let X and Y be independent random variables, with respective probability density functions $f_X(x)$ and $f_Y(y)$. Then, for all values of X and Y

$$f(x,y) = f_X(x)\, f_Y(y) \tag{6}$$

The sum Z = X+Y is a new random variable, and the probability of the event Z=z is defined as the union of all combinations of x and y which sum to z. For continuous functions this relation is given explicitly as

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x)\, f_X(x)\, dx = \int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy \tag{7}$$

which is called the convolution formula [32, p. 186]. Using only the far right hand side of equation (7), the cumulative density function $F_Z(Z^\circ)$ of the sum of X and Y is

$$F_Z(z^\circ) = \int_{-\infty}^{z^\circ} f_Z(z)\, dz = \int_{-\infty}^{z^\circ}\int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy dz \tag{8}$$

For empirical applications with discrete observations, the dimensions of equation (8) can be reduced substantially. If $f_Y(y)=0$ or $f_X(z-y)=0$ then $f_Z(z)=0$ also. This implies that the range of the first integrand can be bounded by the minimum of the ordered y vector and the value of y for which (z-y) falls below the empirical distribution of $f_X(x)$. These values shall be denoted infy and supy, respectively. Similarly the second integral can be bounded from below by the minimum possible value for X+Y, denoted here as infz. In this manner equation (6) can be restated for discrete observations obtained from simulation procedures as[5]

$$\hat{F}_Z(Z^\circ) = \sum_{infz}^{z^\circ} \sum_{infy}^{supy} \hat{f}_X(z-y)\,\hat{f}_Y(y)\,\Delta y \Delta z \tag{9}$$

where $\hat{F}_Z(Z^o)$, $\hat{f}_X(x)$ and $\hat{f}_Y(y)$ are discrete approximations of $F_Z(Z^o)$, $f_X(x)$ and $f_Y(y)$. The incremental values for y and z are defined by the desired level of precision and computational power.

This analysis can be extended to the difference between two empirical distributions (for example two distributions of mean values calculated using either the PLC or the KR methods) by using equation (9) with Y redefined as -Y. Alternatively, a difference variable V= X-Y can be defined such that

$$f_V(v) = f_{X-Y}(v) = \int_{-\infty}^{\infty} f_Y(x-v)\,f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(v+y)\,f_Y(y)\,dy \tag{10}$$

Analogues for equations (8) and (9) are readily derived.

The above equations can be directly applied to the information provided from the full PLC or DP empirical distributions. As in the PLC method, the distribution of the differences will generally not be known, and an empirical approach to estimating confidence intervals is necessary. The creation of $(1-\alpha)$ empirical confidence intervals for the convoluted values is subtly different from the PLC approach. For the percentile approach suggested in PLC and DP each simulated observation has equal probability, and a fixed number of observations ($B^*\alpha/2$ or $B'^*\alpha/2$) is dropped from each tail of the ranked empirical values. In contrast, the convoluted values will not necessarily have equal probabilities of occurrence, and the points of truncation must be determined from the cumulative distribution functions for the convolution. Adopting a 'percentile approach' [16] the lower bound and upper bound of the confidence intervals are respectively defined as

$$\hat{Z}_{low}(\alpha)=\hat{F}_Z^{-1}(\alpha/2) \qquad \hat{Z}_{up}(\alpha)=\hat{F}_Z^{-1}(1-\alpha/2) \tag{11}$$

And,

$$[\hat{Z}_{low}(\alpha),\hat{Z}_{up}(\alpha)] \tag{12}$$

is the approximate $(1-\alpha)$ central confidence interval for Z.

Combining the principle of the two sided difference in means test with a percentile approach, the null hypothesis that the difference between Y and X equals zero is accepted at the α level of significance if the approximate $(1-\alpha)$ confidence interval includes zero and rejected otherwise. Alternatively, assuming that the distributions are ordered in a descending fashion, the approximate significance of the difference between distributions is determined by twice the value of the cumulative distribution function at the convoluted value of zero.

Application to Grand Canyon White Water Rafting

In this section the convolutions method is applied to evaluating the differences between four flow scenarios for commercial white water boaters in the Grand Canyon. In order to focus on the convolutions method, the model presented here is intentionally simplistic - only the bid value is included as an independent variable in the statistical analysis. More sophisticated models and a greater description of the study are presented in Bishop et al. [5], Bishop et al. [6] and Boyle, Welsh and Bishop [10].

The amount of water flowing in the Colorado river is controlled by releases from the Glen Canyon Dam. These flow levels, measured in cubic feet per second (cfs), have a direct and indirect effect on important trip attributes:

> "Time at attraction sights, such as Indian ruins and side canyons with pleasing scenery, and for layovers, depends on the speed of the current. The size and the number of rapids are affected by dam releases. Boaters, particularly those on commercial trips, enjoy fairly large rapids that depend on substantial flows. At relatively low flows and flood flows, passengers, particularly those on commercial oar powered trips, may have to walk around rapids. This is generally considered undesirable by passengers. Flood flows may raise concerns about safety in the minds of boaters. Some risk at rapids makes the trip more exciting, but higher flows (say 40,000 cfs and above) may be perceived to be hazardous by many. The lack of crowding is also important to many boaters. High and flood flows can contribute to crowding at beaches and attraction sites by inundating beaches" [6, p. 11-12]

Apparently, somewhere between the abject terror and congestion of flood waters and the inconvenience and occasional dangers of low flows lies an optimal rate of release that maximizes recreation quality for each boater.

Given the above considerations, the surplus value associated with each flow level can be defined implicitly by

$$\boldsymbol{V}(1, Y-H_j; f_j) = \boldsymbol{V}(0, Y; f_j) \qquad \textbf{(13)}$$

where V(.) is the indirect utility function, Y is income, H_j is the Hicksian compensating surplus (willingness to pay), f_j is the jth flow being evaluated, 1 indicates that the trip is taken, and 0 represents no trip. Because boaters only experience the flow level on their individual trip, it is important to note that the scenarios evaluated here are hypothetical and are based on descriptions provided in the valuation study [10].

Using the linear formulation of the difference function individual logit equations were estimated to predict willingness to pay under four different scenarios -- 5,000 cfs, 13,000 cfs, 22,000 cfs and 40,000 cfs. Each respondent was asked to indicated whether a trip would have been worthwhile at various levels of expense. As presented in Table 1, the estimations are fairly robust. Individual parameters are all significant at the 5 percent level.[6] High χ^2 values for the equations also indicate that each equation is significant at this level.

In addition, as demonstrated in Table 2, each of the estimated equations is significantly different from the estimates for the other three flow levels. Thus, we can conclude that written descriptions of trips taken at different flow levels do significantly affect the distribution of willingness to pay among commercial boaters.

Whether this leads to significant differences in Hicksian surplus is a different question, one that can only be answered by comparing distributions of willingness to pay estimates. Formally the hypothesis being tested is that

$$H_j = H_k \qquad j \neq k \tag{14}$$

where H_j and H_k are the Hicksian surpluses associated with different flow scenarios.

Estimated means and their distributions for each scenario were created using the following closed-form solution of equation (3) [19].

$$E(WTP) = \frac{1}{\beta}\ln(1+e^{\alpha}) \tag{15}$$

where α and β correspond to the coefficients defined in equation (2). In calculating the empirical distributions, intervals for Δy and Δz were set at 1. Critical points on these distributions are presented in Table 3 for both the PLC and DP methods. Evaluation of this table indicates four important aspects of these distributions that merit further discussion. First, as hypothesized, Hicksian surplus appears to be maximized at some intermediate flow level. Second, there is considerable overlap between most of the estimated distributions, which indicates that statistical differences between equations may not necessarily translate into differences in willingness to pay. Third, as demonstrated by the high skewness coefficients in Table 3, most distributions are significantly skewed and apparently deviate from normality. As a result of this latter observation, classical difference tests based on normality assumptions are not relevant here. The only apparent exception to this generalization is found in the DF estimate for flow levels of 40,000 cfs.

The fourth interesting aspect of Table 3 is that, in spite of the difference between methods, the PLC and DF resampling procedures provide remarkably similar estimates of means and medians. In this application the estimates across methods never deviate by more than 2.8 percent for the mean and 1.7 percent for the median. Although the differences between the upper and lower tails vary more widely across estimation methods, the different techniques again provide very similar estimates. On the other hand, it is interesting to note that the DP approach is less skewed across all flow scenarios.

Using the convolutions method detailed in this paper for either the DP or the PLC approach, the null hypothesis is rejected for all pairs except 13 and 40 cfs. As demonstrated in Table 4, this is the only combination of flows for which the approximate 95 percent confidence interval for the difference includes zero. Alternatively stated, α is less than 5 for all other pairwise combinations. For the null hypothesis $H_{13} = H_{40}$ the significance level is approximately 14.59 for the PLC distribution and 13.47 using the DF approach[7].

Thus, the application of either the PLC or the DP estimates results in the same rejection/acceptance conclusions at the 5 percent level in this example. Estimates of the exact significance are also reasonably similar, but do, however, deviate in a systematic manner with the DP approach providing slightly lower estimates of significance across all pairwise comparisons. As with the observation of relative skewness made above, we do not know whether this is an anomaly of the data and estimation techniques used here or is an indication of underlying biases in the two approaches. For the moment, we assert that the deviation in estimates is not large enough to merit concern.

Summary and Conclusions

Dichotomous choice formats are now applied frequently in contingent valuation studies, but previously developed significance tests for differences in empirical distributions of benefit measures have either invoked normality assumptions or used non-overlapping confidence interval criteria. This paper demonstrates that such methods will generally not be appropriate, and develops a exact empirical test, based on the method of convolutions, for assessing the statistical significance between distributions of dichotomous choice contingent valuation welfare estimates. Application of the proposed convolutions approach is illustrated in a case study using two alternative techniques of generating empirical distributions from dichotomous choice data.

This convolutions approach is not limited to the two methods of generating distributions applied in this paper. It is applicable to any method which generates empirical distributions. Moreover, the convolutions formula is not limited to dichotomous choice contingent valuation. The technique is very general and could be applied to any other parameter estimates for which empirical distributions are utilized and comparisons across distributions are desired.

NOTES:

1. Here, the surplus measure is broadly interpreted to include all possible welfare measures suggested in the valuation literature. The intention is to abstract from the important debates over the mean versus the median, equivalent versus compensating variation and surplus, and willingness to pay and willingness to accept approaches.

2. This point has been noted elsewhere for elasticities [14].

3. Duffield and Patterson, citing Smith, Saven and Robertson, and Jennings, note that this assumption may not be valid for small samples.

4. Conversely, equation (5) can be rearranged and solved for the difference between two means that corresponds to a non-overlapping confidence interval for α=0.05. Simple algebra and a critical value of 1.9600 indicate that the point where the two means is significantly different occurs when the means are approximately 2.772 standard errors apart. At this distance, the non-overlapping two-sided confidence intervals only encompass approximately 87 percent of their respective distributions.

5. Appendix 1 demonstrates this technique for simple distributions.

6. Only the intercept term for 5,000 cfs has what could be considered an unexpected sign. A plausible explanation for this negative coefficient is that hicksian surplus may be zero at low flows for many boaters.

7. A different conclusion is reached if we apply the non-overlapping criteria advocated by PLC. Using their method we would conclude that only the estimated Hicksian values 13 and 22 cfs and 13 and 40 cfs are <u>not</u> significantly different at the 5 percent level. The empirical confidence intervals for the differences between all other pairs do not overlap. Based on the previous discussion, this difference in acceptance/ rejection of the null hypothesis $H_{13} = H_{22}$ indicates that the use of the PLC criteria could actually lead to type II errors in policy relevant applications.

TABLE 1
Estimated Logit Equations for Different Flow Scenarios for Commercial White Water Boaters

Conditions[a,b]	5 CFS	13 CFS	22 CFS	40 CFS
Intercept	−0.5516 (−2.1925)	1.7978 (5.8705)	1.9752 (5.8265)	0.9142 (3.3095)
Bid Amount	−0.0025 (−4.1870)	−0.0040 (−6.9497)	−0.0035 (−6.6797)	−0.0030 (−6.3404)
Model χ^2	28.1843	77.3268	60.7446	58.7822
McFadden R^2	0.11	0.19	0.15	0.16
N	297	297	297	297

[a] Asymtotic t-values in parentheses.
[b] $t_{0.05,295} = 1.9600$.

TABLE 2
χ^2 Values between Estimated Logit Equations for Different Flow Scenarios for Commercial White Water Boaters[a]

	13 CFS	22 CFS	40 CFS
5 CFS	74.84	96.89	34.57
13 CFS		6.03	6.00
22 CFS			18.10

[a] $\chi^2_{0.05,2} = 5.99$.

TABLE 3
Empirical Willingnes to Pay for Different Flow Scenarios Based on 1000 Draws

Flow in 1,000s	Method	Calcualated from Parameter Means	Based on 1000 Draws				
			Mean	Lower Tail 5%	Median	Upper Tail 5%	Skewness
5 CFS	KR/PLC DP	181	189 184	139 135	184 181	261 252	0.99 0.59
13 CFS	KR/PLC DP	485	489 488	433 428	486 486	558 553	0.45 0.21
22 CFS	KR/PLC DP	604	607 605	543 538	605 604	680 681	0.45 0.27
40 CFS	KR/PLC DP	415	419 416	353 346	417 415	497 488	0.40 0.07

[a] $g_{0.01,1000} = 0.180$ [Table 34b; Tables for Statisticians and Biometricians].

[b] KR/PLC refers to the Krinsky and Robb and Park, Loomis and Creel Technique. DP refers to the "bootstrap" method developed in Duffield and Patterson.

TABLE 4
Approximate 95 Percent Conidence Intervals and Significance Levels between Mean Willingness to Pay Estimates for Different Flow Scenarios for Commercial White Water Boaters Using KR/PLC Empirical Distributions[a,b]

	13 CFS	22 CFS	40 CFS
5 CFS	[207,387] 0.0000	[321,508] 0.0000	[130,324] 0.2180
13 CFS		[23,211] 1.6526	[-27,167] 14.5910
22 CFS			[87,288] 0.0824

[a] The set of numbers in the brackets corresponds to the approximate 95 percent confidence interval associated with equation (10) in the text.

[b] The numbers below the brackets indicates the approximate two-sided significance level (α) evaluated at 0. If $\alpha < 5$ then the estimates are significantly different at the 5 percent level.

TABLE 5
Approximate 95 Percent Conidence Intervals and Significance Levels between Mean Willingness to Pay Estimates for Different Flow Scenarios for Commercial White Water Boaters Using DP Empirical Distributions[a,b]

	13 CFS	22 CFS	40 CFS
5 CFS	[215,387] 0.0000	[330,511] 0.0000	[139,321] 0.0000
13 CFS		[25,210] 1.2088	[-22,163] 13.4678
22 CFS			[93,287] 0.0002

[a] The set of numbers in the brackets corresponds to the approximate 95 percent confidence interval associated with equation (10) in the text.

[b] The numbers below the brackets indicates the approximate two-sided significance level (α) evaluated at 0. If $\alpha < 5$ then the estimates are significantly different at the 5 percent level.

APPENDIX 1
Demonstration of the Convolutions Approach using Simple Distributions

Using simple distributions, this appendix provides an demonstration of the discrete convolution formula presented in equation (9) and the suggested statistical test for estimating the significance of the difference between two simple empirical distributions. Recall that equation (9) was specified as,

$$\hat{F}_Z(Z^{\circ}) = \sum_{infz}^{z^{\circ}} \sum_{infy}^{supy} \hat{f}_X(z-y)\, \hat{f}_Y(y)\, \Delta y \Delta z \tag{9'}$$

where all parameters are defined in the text.

Suppose that we are interested in evaluating the difference between the two following empirical distributions

.	$f_X(\cdot)$	$f_Y(\cdot)$
0	0	.05
1	0	.3
2	.1	.6
3	.4	.05
4	.4	0
5	.1	0

As noted in the text, equation (9), which is the additive form, can be used to generate a distribution of the difference between two empirical distributions by redefining Y as -Y. This redefinition yields

.	$f_X(\cdot)$	$f_Y(\cdot)$
-3	0	.05
-2	0	.6
-1	0	.3
0	0	.05
1	0	0
2	.1	0
3	.4	0
4	.4	0
5	.1	0

APPENDIX 1 (cont.)

With this redefinition, the convolution formula results in the following distributions of the difference between the two distributions.

			$F_Z(-2) = .000$
$f_Z(-1) =$	$f_x(2)f_y(-3)$ $(1)(.05)$	$= .005$	$F_Z(-1) = .005$
$f_Z(0) =$	$f_X(2)f_Y(-2) + f_X(3)f_Y(-3)$ $(.1)(.6) + (.4)(.05)$	$= .080$	$F_Z(0) = .085$
$f_Z(1) =$	$f_X(2)f_Y(-1) + f_X(3)f_Y(-2) + f_X(4)f_Y(-3)$ $(.1)(.3) + (.4)(.6) + (.4)(.05)$	$= .290$	$F_Z(1) = .375$
$f_Z(2) =$	$f_X(2)f_Y(0) + f_X(3)f_Y(-1) + f_X(4)f_Y(-2) + f_X(5)f_Y(-3)$ $(.1)(.05) + (.4)(.3) + (.4)(.6) + (.1)(.05)$	$= .370$	$F_Z(2) = .745$
$f_Z(3) =$	$f_X(3)f_Y(0) + f_X(4)f_Y(-1) + f_X(5)f_Y(-2)$ $(.4)(.05) + (.4)(.3) + (.1)(.6)$	$= .200$	$F_Z(3) = .945$
$f_Z(4) =$	$f_X(4)f_Y(0) + f_X(5)f_Y(-1)$ $(.4)(.05) + (.1)(.3)$	$= .050$	$F_Z(4) = .995$
$f_Z(5) =$	$f_X(5)f_Y(0)$ $(.1)(.05)$	$= .005$	$F_Z(5) = 1.000$

The significance of the differences between distributions is 0.085, as demonstrated by $F_Z(0)$ on the above table.

APPENDIX 2
Comparison of Convolutions Formula and Analytical Difference Between Means

For validation purposes, this appendix demonstrates that the convolutions approach closely approximates the analytical solution for the differences in means. Recall that, for distributions with equal variance, the analytical solution for the difference in means is given by:

$$Z=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{2}*(\sigma/\sqrt{n})}=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{2}*STDERR} \qquad \sim N(0,1)$$

where z is the test statistic, X_i are the sample means, n is the number of observations, σ is the standard deviation of the distribution, and STDERR is the associated standard error of the mean [Snedecor and Cochrane, 101].

Suppose now that we are interested in comparing the means of two distributions with the following characteristics

	n	X	σ	STDERR
Y	1000	12	10	0.316
X	1000	12.75	10	0.316

Using the analytical formula presented above, yields

$$Z=\frac{(12.75-12)}{\sqrt{2}*0.316}\approx 1.677$$

The test statistic 1.677 corresponds to a two tailed significance level of 0.0936.

Alternatively, using the discrete convolutions formula presented in equation (9) with increments set at 0.01 provides a two tailed significance level of 0.098. The small difference between the analytical significance and the empirical significance calculated here is due to the random nature of the draws and the discreetness of the approach.

References:

1. Adamowitz, W. L., J. L. Fletcher, and T. Graham-Tomasi, 1989. "Functional Form and Statistical Properties of Welfare Measures", American Journal of Agricultural Economics, 71:414-421.

2. Bergstrom, J. C., J. R. Stoll and A. Randall, 1990. "The Impact of Information on Environmental Valuation Decisions", American Journal of Agricultural Economics, 72:614-21.

3. Bishop, R. C. and T. A. Heberlein, 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?", American Journal of Agricultural Economics, 61:926-930.

4. Bishop, R. C., T. A. Heberlein, and M. J. Kealy, "Contingent Valuation of Environmental Assets: Comparisons with a Simulated Market", Natural Resources Journal, 23(2):619-33.

5. Bishop, R. C., et al., 1987. "Glen Canyon Dam Release and Downstream Recreation: An Analysis of User Preferences and Economic Values", Final Report to the U.S. Bureau of Reclamation, Salt Lake City.

6. Bishop, R. C., et al., 1989. "Grand Canyon Recreation and Glen Canyon Dam Operations: An Economic Evaluation", Paper Presented at the joint meetings of the Western Committee on the Benefits and Costs of Natural Resource Planning (W133) and the Western Regional Science Association, San Diego, CA.

7. Bowker, J. M., and J. M. Stoll, 1988. "Dichotomous Choice in Resource Valuation", American Journal of Agricultural Economics, 70(2):372-382.

8. Boyle, K. J., 1989. "Commodity Specification and the Framing of Contingent-Valuation Questions", Land Economics, 65:57-63.

9. Boyle, K. J., and R. C. Bishop, 1988. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques", American Journal of Agricultural Economics, 70(1):20-28.

10. Boyle, K. J., M. P. Welsh, and R. C. Bishop, 1988. "Validation of Empirical Measures of Welfare Change: Comment, Land Economics, 64(1):94-98.

11. Boyle, K. J., M. P. Welsh and R. C. Bishop, 1991. "An Empirical Investigation of the Role of Question Order, Symbolic Effects, and Respondent Experience in Contingent-Valuation Studies", Unpublished Manuscript.

12. Cameron, T. A., 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censured Logistic Regression", Journal of Environmental Economics and Management, 15:355-379.

13. Cameron, T. A., 1991. "Interval Estimates of Non-Market Resource Values from Reverendum Contingent Valuation Surveys", Land Economics, 67:413-21.

14. Dorfman J. H., C. L. Kling and R. J. Sexton, 1990. "Confidence Intervals for Elasticities and Flexibilities: Reevaluating the Ratios of Normals Case", American Journal of Agricultural Economics, 72:1006-1017.

15. Duffield, J. W. and D. A. Patterson, 1991. "Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation" Land Economics, 67(2):225-239.

16. Effron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans, Monograph 38, Society Industrial and Applied Mathematics, Philadelphia.

17. Feller, W., 1957. An Introduction to Probability Theory and Its Application, Volume I, 2nd edition, John Wiley and Sons, Inc., New York.

18. Hanemann, W. M., 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses", American Journal of Agricultural Economics, 66(3):332-341.

19. Hanemann, W. M., 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses: Reply", American Journal of Agricultural Economics, 71(4):1057-1061.

20. Hoehn, J. P. and A. Randall, 1987. "A Satisfactory Benefit-Cost Indicator from Contingent Valuation", Journal of Environmental Economics and Management, 14(3):226-247.

21. Jennings, D., 1986. "Judging Inference Adequacy In Logit Regression", Journal of the American Statistical Association, 81:471-476.

22. Johnson, P-O, B. Kriström, K. G. Mäler, 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data:Comment", American Journal of Agricultural Economics, 71:1054-1056.

23. Kahneman, D. and J. L. Knetsch, 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction", Journal of Environmental Economics and Management, 22: 57-70.

24. Kealy, M.J., M. Montgomery and J. P. Dovidio, 1990. Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter", Journal of Environmental Economics and Management, 19:244-263.

25. Kling, C. L., 1991. "Estimating the Precision of Welfare Measures". Journal of Environmental Economics and Management, 21(3):244-259.

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

27. Kriström, B., 1990. "A Non-parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Response Studies", Land Economics, 66:135-139.

28. Loomis, J. B., 1989. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses", American Journal of Agricultural Economics, 71:76-84.

29. Loomis, J. B., 1987. "Balancing Public Trust Resources of Mono Lake and Los Angeles' Water Right: An Economic Approach", Water Resources Research, 23:23-29.

30. McConnell, K.E., 1990. "models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation", Journal of Environmental Economics and Management, 18:19-34.

31. Mitchell, R. C. and R. T. Carson, 1989. Using Surveys to Value Public Goods: The Contingent Valuation Approach, Resources for the Future, Washington, D.C.

32. Mood, A., F. Graybill, and D. Boes, 1974. Introduction to the Theory of Statistics, McGraw Hill, San Francisco.

33. Park, T., J. B. Loomis and M. Creel, 1991. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies", Land Economics, 67(1):64-73.

34. Pearson, E. S. and H. D. Hartley, eds., 1966. Biometrika Tables for Statisticians (third edition), University Press, Cambridge.

35. Smith, K. C. , N. E. Savin, and J. L. Robertson, 1984. "A Monte Carlo Comparison of Maximum Likelihood and Minimum Chi Square Sampling Distributions in Logit Analysis", Biometrics, 40:471-482.

36. Snedecor, G. W. and W. G. Cochrane, 1980. Statistical Methods, 7th ed. Iowa State University Press, Ames.