Measuring the Difference in Mean Willingness to Pay When Dichotomous Choice Contingent Valuation Responses Are Not Independent

Gregory L. Poe, Michael P. Welsh, and Patricia A. Champ

ABSTRACT. *Dichotomous choice contingent valuation surveys frequently elicit multiple values in a single questionnaire. If individual responses are correlated across scenarios, the standard approach of estimating willingness to pay (WTP) functions independently for each scenario may result in biased estimates of the significance of the difference in mean WTP values. This paper applies an alternative bivariate probit approach that explicitly accounts for correlation across errors in the estimation of WTP and mean WTP. Using data from three separate dichotomous choice contingent valuation studies, this correlation is demonstrated to have an effect on the significance level of mean WTP difference tests.* (JEL Q26)

I. INTRODUCTION

The high costs associated with collecting primary data, coupled with a need for within-sample comparisons of Hicksian surplus values, frequently lead researchers to include several contingent valuation (CV) questions in a single survey. In particular, resource valuation surveys often elicit surplus values for a baseline level of resource provision (e.g., current hunting conditions) and then ask about values for alternative levels of provision (e.g., improved hunting conditions).[1] Although this approach reduces data collection costs and allows for the estimation of continuous resource valuation functions (e.g., Boyle, Welsh, and Bishop 1993), possible correlation between responses complicates policy relevant comparisons of expected benefits across scenarios. Such complications similarly arise in testing for within-subject embedding effects, in which values placed on a comprehensive good are compared with values for a subset of the comprehensive good (e.g., Carson and Mitchell 1995).

Correlation across valuation response functions, or more formally across errors, will be associated with the extent that estimated models fail to capture individual specific factors that have a common effect on responses across questions. Although it is recognized that individual valuation processes may be complex and heterogeneous, most estimated valuation functions consist of relatively simple models in which the combined effects of excluded variables are assumed to be summarized by a random disturbance. If error terms include systematic components, unmeasurable or omitted variables that represent factors particular to individuals are likely to create correlation in estimated errors across equations (Hsiao 1986). The direction of correlation should be affected by the perceived likeness of the goods being valued: closely related, embedded, or nested goods would likely result in a positive correlation in error terms; goods with divergent bundles of characteristics may exhibit positive or negative correlation, or independence, depending on the nature of

Poe is with the Department of Agricultural, Resource, and Managerial Economics, Cornell University; Welsh is with Hagler Bailly Consulting, Inc., Madison, WI; and Champ is with the Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO. The authors are indebted to Bill Provencher and Jean-Paul Chavas for raising the independence question in the context of another paper. Richard Bishop provided helpful comments and was instrumental in the funding, design, and implementation of the surveys used in this study. Tim Mount, Carlos Reberte, and Joseph Cooper provided helpful programming advice. Participants in the 1996 USDA Regional Project W-133 meetings and two anonymous reviewers provided useful comments. Of course, none of the above are responsible for any errors in this manuscript. Funding for this project was provided by the College of Agriculture and Life Sciences, Cornell University; College of Agricultural and Life Sciences, University of Wisconsin–Madison; and the Glen Canyon Environmental Studies Project, U.S. Bureau of Reclamation.

[1] Papers by Park, Loomis, and Creel (1991), and Boyle, Welsh, and Bishop (1991, 1993) provide examples of this format.

Land Economics • May 1997 • 73 (2): 255–67

the alternative sets of characteristics. Correlated responses could also be caused by systematic response patterns associated with CV such as "yea saying" (Kanninen 1995), "symbolic" effects (Boyle, Welsh, and Bishop 1991), "warm glow" and "embedding" effects (Kahneman and Knetsch 1992), or "starting point" biases (Cameron and Quiggin 1994). In all, many factors are relegated to the error term in analyses of CV responses, increasing the likelihood that errors will be correlated across valuation response categories.

To the extent that individual responses to successive scenarios within a survey are correlated, the standard CV approach of estimating independent willingness-to-pay (WTP) functions will provide biased estimates of the variance of mean WTP. In turn, the significance of the difference of estimated mean WTP values between scenarios will be biased. Two factors underlying this bias are most easily distinguished by referring to the well-known formula for the variance of the difference of two normal distributions: $\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2^*\text{cov}(X,Y)$. The first factor affecting this difference is an "efficiency effect" associated with the estimation of individual distributions. Although point estimates from independently estimated WTP models remain consistent even if responses are correlated across scenarios, the estimates are inefficient (Fahrmeir and Tutz 1994). Because estimated distributions of mean WTP, depicted here as X and Y, are derived from coefficients of the estimated WTP functions, the dispersion of estimated individual mean WTP distributions will subsequently be biased if the analyst does not account for correlation. There may also be efficiency gains associated with imposing restrictions across equations.[2] The second factor affecting the distribution of the difference is a "correlation effect" in that the failure to account for correlation between X and Y, depicted above by the covariance, will lead to biased estimates of the variance of the difference between these two variables.

Using data from three dichotomous choice CV resource studies as examples, this paper investigates the impact that the efficiency effect has on the estimates of individual WTP distributions, and the combined impact of both the efficiency and the correlation effects on the variance and the significance of the difference of the estimated mean WTP distributions. The remainder of the paper is organized as follows. Section II provides the conceptual framework for the bivariate probit and bootstrapping approaches used in the analysis. The CV studies used to investigate these issues are described in Section III. Empirical results are discussed in Section IV, and Section V provides the conclusions from this study and implications for future research.

II. CONCEPTUAL FRAMEWORK

Assume that the ith individual has some true surplus value (s_{ji}) for the good described in the jth scenario, and that the respondent will indicate ($I_{ji} = 1$) that they are willing to pay the posted price (p_{ji}) if $s_{ji} \geq p_{ji}$. If $s_{ji} < p_{ji}$ the individual will not be willing to pay p_{ji}, and $I_{ji} = 0$. Following the random utility framework presented in Hanemann (1984) and Cameron and coauthors (Cameron and James 1987; Cameron and Quiggin 1994), assume that the unobserved value $s_{ji} = \beta_j' x_{ji} + u_{ji}$, where the systematic component, $\beta_j' x_{ji}$, is a function of a vector, x_j, of observable attributes of the respondent, including the dichotomous choice posted price, and u_j is an unobservable random disturbance assumed to be distributed $N(0, \sigma_j^2)$.

Standard approaches to evaluating and testing alternative scenarios assume that the

[2] This point was raised by an anonymous reviewer. The reader is referred to a related paper by Alberini and Kanninen (1994) which explores the efficiency gains associated with joint estimation and cross-equation restrictions of various combinations of continuous and discrete contingent valuation response formats.

u_{ij} are uncorrelated across scenarios.[3] Under this assumption WTP distributions are estimated independently for each scenario. Approximate distributions of estimated mean WTP are derived from these estimated functions by applying bootstrapping or other repeated sampling techniques (Park, Loomis, and Creel 1991). Under the independence assumption, the difference in approximate distributions of estimated mean WTP can be estimated using an empirical convolutions approach or by directly bootstrapping the difference (Poe, Severance-Lossin, and Welsh 1994).[4]

As suggested previously, the assumption of independence may not be appropriate if multiple CV questions are posed in the same questionnaire. Econometrically, in a manner analogous to seemingly unrelated regressions, this non-independence between the two valuation functions for scenarios one (s1) and two (s2) may be accommodated by explicitly accounting for cross-equation correlation in the estimation process. Within a discrete choice format, this can be accomplished by assuming a bivariate normal distribution $\mathrm{BVN}(\beta_1'x_1, \beta_2'x_2, \sigma_1^2, \sigma_2^2, \rho)$ of the errors where β_j and x_j correspond to parameters previously defined and ρ is the correlation coefficient. Defining $z_1 = -\beta_1'x_1/\sigma_1$ and $z_2 = -\beta_2'x_2/\sigma_2$ to be standardized normal errors, the standard bivariate normal distribution $\mathrm{SBVN}(\rho)$ for (z_1, z_2) takes the following form.

$$\Phi(z_1, z_2; \rho) = \frac{\exp^{(z_1^2 - z_2^2 - 2\rho z_1 z_2)/2(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}}. \qquad [1]$$

If ρ is indeed zero, this density function collapses to the product of two independent normal density functions, and the univariate approach outlined previously is appropriate for estimating separate probit WTP distributions, independently estimated mean WTP distributions, and the difference of the estimated mean WTP distributions. If $\rho \neq 0$, the associated likelihood function for the four possible pairs of responses [yes (s1)–yes (s2), yes (s1)–no (s2), no (s1)–yes (s2), no (s1)–no (s2)] across equations is given as:

$$L = \prod_i \left[\int_{\frac{\beta_1'x_1}{\sigma_1}}^{\infty} \int_{\frac{\beta_2'x_2}{\sigma_2}}^{\infty} \Phi(z_1, z_2; \rho)\, dz_1\, dz_2 \right]^{I_1 I_2} *$$

$$\left[\int_{\frac{\beta_1'x_1}{\sigma_1}}^{\infty} \int_{-\infty}^{\frac{-\beta_2'x_2}{\sigma_2}} \Phi(z_1, z_2; \rho)\, dz_1\, dz_2 \right]^{I_1(1-I_2)} *$$

[3] The motivation and conceptual framework parallels that used to support bivariate probit models in double bounded dichotomous choice contingent valuation detailed in Cameron and Quiggin (1994) and Alberini (1995). Hoehn and Loomis (1993) offer an alternative modeling approach that explicitly accounts for substitution across dichotomous choice scenarios rather than relegating these effects to the error term. However, empirical implementation of their approach requires a very structured data set exceeding the design of most dichotomous choice surveys that ask multiple valuation questions. As presented, the Hoehn and Loomis approach is particularly suited to evaluating discrete, independent policy packages, and further assumes that errors in modeling responses are independent. Future research might seek to synthesize the Hoehn and Loomis approaches with the bivariate approach used in this paper by explicitly accounting for cross-scenario substitution effects while allowing for correlation among the error terms.

[4] Beginning with an original data set consisting of N observations, bootstrapping randomly samples, with replacement, N observations to create an artificial data set from which coefficients and mean WTP are estimated (see Efron and Tibshirani 1993). This resampling is repeated a large number (M) times to create distributions of the estimated coefficients and derived mean WTP values. A second resampling procedure, used widely in dichotomous choice contingent valuation and this study, randomly draws M simulated coefficient values from the maximum likelihood estimates and the associated covariance matrix (see Krinsky and Robb 1986). Mean WTP estimates are calculated for each of these M draws creating an empirical distribution of estimated mean WTP. Past approaches for comparing empirical distributions such as the non-overlapping confidence interval criterion (e.g., Park, Loomis, and Creel 1991) or normality assumptions (e.g., Desvousges et al. 1992) are biased or are otherwise not appropriate for general applications (Poe, Severance-Lossin, and Welsh 1994). Moreover, empirical bootstrapping approaches have been shown to approximate analytical solutions in general (Efron and Tibshirani 1993) and for dichotomous choice contingent valuation in particular (Balistreri et al. 1996).

$$\left[\int_{-\infty}^{\frac{-\beta_1' x_1}{\sigma_1}} \int_{\frac{-\beta_2' x_2}{\sigma_2}}^{\infty} \Phi(z_1, z_2; \rho)\, dz_1\, dz_2\right]^{(1-I_1)I_2} *$$

$$\left[\int_{-\infty}^{\frac{-\beta_1' x_1}{\sigma_1}} \int_{-\infty}^{\frac{-\beta_2' x_2}{\sigma_2}} \Phi(z_1, z_2; \rho)\, dz_1\, dz_2\right]^{(1-I_1)(1-I_2)} . \qquad [2]$$

The hypothesis H_0: $\rho = 0$ can be evaluated with a standard likelihood ratio test, $-2(\mathrm{LL}_1 + \mathrm{LL}_2 - \mathrm{LL}_j) \sim \chi^2$, by comparing the log of this likelihood function (LL_j) with the sum of the log likelihoods ($\mathrm{LL}_1, \mathrm{LL}_2$) associated with the independently estimated probit distributions (Greene 1993). Similarly, a comparison of log likelihood values can be used to assess the validity of cross-equation restrictions on the estimated parameters. In making such comparisons, it should be noted that the greatest efficiency gains are expected when $X_1 \neq X_2$. But, in contrast to continuous dependent variables, there should also be efficiency gains even when the covariates are identical across equations (Alberini and Kanninen 1994).

If the null hypothesis H_0: $\rho = 0$ is rejected, $E(WTP_2|WTP_1)$ is a non-zero function of ρ (Goldberger 1991) and the mean WTP distributions will depend upon the joint distribution of estimated parameters, one of which is ρ. Consequently, simulated mean WTP distribution values from the joint distribution must be paired, and the difference of the estimated mean WTP distributions $\overline{WTP_{jB}}$ can be estimated by directly bootstrapping the difference,

$$D = \overline{WTP_{1m}} - \overline{WTP_{2m}} \qquad m = 1, \ldots, B \qquad [3]$$

where B is the number of paired bootstrap observations. Following the percentile approach in Efron and Tibshirani (1993), the approximate one-sided significance of the difference is obtained by computing the proportion of negative values in D.

III. DATA

The data for this analysis were taken from three separate dichotomous choice CV mail surveys of recreational resource use. Examples of individual CV questions from each of the surveys are provided in the Appendix.

The *Escanaba Lake Survey* was conducted as part of a study to assess the validity of CV values by comparing hypothetical WTP to actual WTP.[5] Escanaba Lake is one of five lakes managed by the Wisconsin Department of Natural Resources in the Northern Highland State Forest of Vilas County. It is the only lake in Northern Wisconsin where anglers can fish for walleye after the ice is off the lake before the regular fishing season. This early season between "ice-off" and the regular fishing season can vary from a few days to a few weeks. Individuals who had fished the early season at Escanaba Lake in 1989, 1990, or 1991 were mailed a CV questionnaire in March of 1992 (prior to the early season). Eight hundred and twenty questionnaires were mailed and 621 were completed. Adjusting for undeliverable questionnaires, the response rate was 82 percent. The questionnaire included two dichotomous choice CV questions. The first question asked whether the individual would pay $X for a "baseline permit" to fish the upcoming early season, in which expected catch corresponds to historical levels. The second CV question asked whether the respondent would pay $Y for a permit to fish the upcoming early season if there would be "15 percent fewer" walleye than usual in Escanaba Lake. The format of these questions is that the second CV question ("15 percent fewer") is nested in the baseline case, in that, with exception of the number of fish available, the scenarios are identical.

The 1991 and 1992 *Sandhill Public Deer Hunt Surveys* were part of a larger study to assess the ability of recreationists to recall expenses related to a special deer hunt (see

[5] Unfortunately, the actual WTP data were never collected.

Champ and Bishop 1996 for further details). Sandhill Wildlife Demonstration Area is a wildlife research property managed by the Wisconsin Department of Natural Resources in Wood County, Wisconsin. In 1991, 352 one-day deer hunting permits were issued for an either sex deer hunt to be held in November. One hundred seventy-seven of the permit holders were sent questionnaires after the hunt. Seventy of the permit holders who were sent a questionnaire did not attend the hunt at Sandhill. Of the 107 hunters who received a questionnaire and hunted, 104 (97 percent) returned the questionnaire. In 1992, the November Sandhill hunt was for antlerless deer only. Two hundred thirty permits were issued and 117 hunters were sent a questionnaire. One hundred seven (91 percent of the deliverable questionnaires) questionnaires were returned. The questionnaires sent in 1991 and 1992 were very similar. Respondents were asked about their expenses related to the Sandhill deer hunt, the quality of the hunt, some demographic questions, and two dichotomous choice CV questions. One CV question asked about their willingness to pay for an "either sex" deer hunting permit and the other asked about their willingness to pay for an "antlerless" deer hunting permit at Sandhill. As with the *Escanaba Lake Survey*, the "antlerless" deer permit is formally a nested subset of the "either sex" permit. However, since each permit is good for only one animal, an element of choice arises. Hunters with "either sex" permits typically report that they do not want to "waste" their permit on does and immature animals, and therefore these permits may be viewed as having non-inclusive elements. *A priori* this trade-off might be expected to have a negative impact on the correlation coefficient.

The objective of the *Grand Canyon White Water Boater Survey* was to estimate a statistical relationship between Hicksian surplus values for white-water trips and average daily Colorado River flows between 5,000 and 40,000 cubic feet per second (see Boyle, Welsh, and Bishop 1993 for further details). In this survey individual respondents were each asked four dichotomous choice CV questions corresponding to the following hypothetical flow levels: 5, 13, 22, and 40 thousand cubic feet per second (kcfs). Prior to answering the valuation questions, respondents answered a series of questions about the attributes of their Grand Canyon white-water trip, including trip expenditures. Each of the valuation questions were preceded by a description of the boating and camping conditions associated with that specific flow. Conducted in 1986, 169 usable responses were obtained from private boaters, representing approximately 91 percent of deliverable surveys. In contrast to the fishing and hunting surveys, each flow level is associated with distinct characteristics, and one flow level cannot be viewed as a nested subset of other flow levels. However, some attributes associated with different flow levels are common even in paired scenarios that describe substantially different flows. For example, both 5 kcfs and 40 kcfs entail inconvenient portaging around additional rapids. Similarly, adjacent flow levels have trip attributes that overlap considerably, but maintain some distinct elements. To the extent that individuals have preferences over flow characteristics, some flow levels might be regarded as substitutes.

IV. RESULTS

Estimated CV responses functions and associated mean WTP values in each of the three surveys were compared with values obtained from different scenarios in the same questionnaire. The procedure for evaluating the effects of cross-scenario correlation was to analyze each pair of questions as follows. First, bivariate (joint) and univariate (independent) probit models were estimated using maximum likelihood techniques. Likelihood ratio tests were used to evaluate the hypothesis that H_0^1: $\rho = 0$ as well as to test various cross-equation equality restrictions. For comparisons in which H_0^1 is not rejected, no additional analyses were conducted beyond the initial maximum likelihood estimates. In the cases where H_0^1 is rejected, 10,000 simulated values of mean WTP were estimated using numerical integration techniques over the non-negative

TABLE 1
VARIABLE DEFINITIONS AND DESCRIPTIVE STATISTICS FOR ESCANABA AND SANDHILL STUDIES

	Description	Mean (s.d.)
Escanaba		
Import	Categorical response "If I could not go fishing at Escanaba Lake during the 'early season,' I would:" 1) easily find something else to do; 2) miss it, but not as much as other things that I enjoy; 3) miss it more than the other interests I now have; 4) miss it more than all the other interests I now have	2.01 (0.89)
Miles	Open-ended variable: distance between Escanaba Lake and home (one way)	120.41 (115.05)
Education	Categorical response: 1) less than high school; 2) high school graduate; 3) some college or technical school; 4) technical or trade school graduate; 5) college graduate; 6) advanced degree	3.20 (1.47)
Bid 1	Dichotomous choice bid value for "baseline permit"	13.60 (12.43)
Bid 2	Dichotomous choice bid value for "15 percent fewer"	11.34 (9.62)
Sandhill		
Quality	Categorical response for quality of the hunt: 1) very low quality; 2) fairly low quality; 3) average quality; 4) fairly high quality; 5) very high quality	3.40 (1.23)
Year	Binary variable: 1991 = 1, 1992 = 2	1.51 (0.50)
Bid 1	Dichotomous choice bid value for "either sex" permit	27.91 (21.81)
Bid 2	Dichotomous choice bid value for "antlerless" permit	29.12 (24.58)

range of the WTP distributions (Hanemann 1984, 1989) and a parametric bootstrap technique that draws simulated coefficient values from the covariance matrix (Park, Loomis, and Creel 1991; Krinsky and Robb 1986). For the jointly estimated bivariate model, the distributions of estimated mean WTP for each question were approximated after accounting for ρ in the estimated covariance matrix. In both the joint and independent models, pairwise differences were calculated as in equation [3]. Comparisons of these approximate distributions of the difference for the joint and independent estimates provide the basis for assessing the effects of the independence assumption on the distribution of the difference. The approximate one-sided significance of the difference is calculated by the proportion of negative values in the distribution of the difference.

The results of this sequence of procedures for the three separate studies are summarized in Tables 1 to 6. Attention in the analyses of efficiency effects is focused, however, on the Escanaba and the Sandhill studies, as they adequately demonstrate the various effects of joint estimation and cross-equation restrictions. Descriptive statistics and definitions of the variables used in the maximum likelihood estimates of the univariate and bivariate probit models for these studies are provided in Table 1. Following Boyle, Welsh, and Bishop (1993), analyses of the *Grand Canyon White Water Boaters* survey responses involved simple models with the only covariates being the cost of the actual trip taken and the bid value for the hypothetical flow scenario.[6]

Table 2 summarizes the independent and joint estimation results for the "baseline permit" and the "15 percent fewer" valuation questions asked in the Escanaba fishing study. The first column presents independently estimated valuation functions. The

[6] An appendix of the WTP distribution estimates of the *Grand Canyon White Water Boaters Survey* is available from the authors.

TABLE 2
ESCANABA FISHING STUDY

	Independent	Joint, Unrestricted	Joint, Restricted
Full Permit			
Constant	−0.5860	−0.5090	−0.5100
	(0.2492)**	(0.2420)**	(0.2335)**
Import	0.2172	0.2488	0.2609
	(0.0777)***	(0.0767)***	(0.0656)***
Miles 1	0.0014	0.0016	0.0013
	(0.0006)**	(0.0006)**	(0.0005)***
Educ	0.1356	0.1239	0.1286
	(0.0489)***	(0.0508)**	(0.0483)***
Bid 1	−0.1229	−0.1350	−0.1371
	(0.0131)***	(0.0100)***	(0.0079)***
15 Percent Fewer			
Constant	−0.3371	−0.3166	−0.3082
	(0.2522)	(0.2573)	(0.2344)
Import	0.2685	0.2738	See Import,
	(0.0760)***	(0.0780)***	full permit
Miles 1	0.0012	0.0011	See Miles 1,
	(0.0006)**	(0.0006)*	full permit
Educ	0.0208	0.0241	0.0186
	(0.0466)	(0.0474)	(0.0472)
Bid 2	−0.1346	−0.1390	See Bid 1,
	(0.0141)***	(0.0126)***	full permit
ρ		0.9173	0.9160
		(0.0426)***	(0.0354)***
Likelihood ratio χ_1^2	180.79		
Likelihood ratio χ_2^2	142.32		
− Log Likelihood[a]	−218.63−228.02	−391.66	−391.94
n	540	540	540

Note: Numbers in ( ) are asymptotic standard errors.
*, **, *** indicate significance levels of 0.10, 0.05, and 0.01, respectively.
[a] $-2(\mathrm{LL}_i - \mathrm{LL}_{j,u}) = 110.04$, $\chi^2_{1,0.10} = 2.71$.

second and third columns present the joint unrestricted and restricted models, respectively. Although varying in significance, the signs of the estimated coefficients are consistent across equations: the probability of a "yes" response increases with perceived importance of the resource, distance traveled to Lake Escanaba, and the educational level of the respondent, but falls with increasing bid values. Importantly, there is an extremely high correlation ($\rho \simeq 0.92$) in estimated response functions across the two dichotomous choice CV questions. Likelihood ratio tests demonstrate that this correlation coefficient is highly significant. Casual comparison of estimated parameters suggests that the WTP response functions are quite similar across scenarios.

A likelihood ratio test of H_0^2: $\beta_{\text{Baseline Permit}} = \beta_{\text{15 Percent Fewer}}$ is rejected at the 5 percent significance level (LR = 11.36, $\chi^2_{5,0.05} = 11.05$), implying that the response functions are different in spite of the fact that they are significantly and highly correlated. The final restricted model was arrived at by testing various individual and joint cross-equation restrictions for coefficients. For this data set, the equality restrictions hold for the estimated coefficients for the Import, Miles 1, and Bid variables. The hypothesis of cross-equation equality for the Education coefficient was rejected. In the baseline scenario equation this coefficient was positive and significant, but was not significant for the 15 percent fewer model. The cause of this difference across equations is not identified.

Inspection of the asymptotic standard errors in each of the models indicates that there is little efficiency gain from estimating

TABLE 3
SANDHILL DEER HUNTING, 1991–1992

	Independent	Joint, Unrestricted
Either Sex		
Constant	0.9977	0.9919
	(0.4245)**	(0.4349)**
Quality	0.1224	0.1278
	(0.0871)	(0.0876)
Year	0.1361	0.1266
	(0.2052)	(0.2074)
Bid 1	−0.0369	−0.0371
	(0.0056)***	(0.0057)***
Antlerless		
Constant	−1.378	−0.1008
	(0.3919)	(0.3941)
Quality	0.2321	0.2288
	(0.0922)**	(0.0933)**
Year	1.1225	1.1245
	(0.2197)***	(0.2323)***
Bid 2	−0.0704	−0.0732
	(0.0129)***	(0.0118)***
ρ		0.3923
		(0.1382)***
Likelihood ratio χ_1^2	60.52	
Likelihood ratio χ_2^2	82.73	
−Log Likelihood[a]	−105.98–93.32	−195.47
n	197	197

Note: Numbers in ( ) are asymptotic standard errors.
*, **, *** indicate significance levels of 0.10, 0.05, 0.01, respectively.
[a] $-2(LL_i - LL_{j,u}) = 7.68$, $\chi^2_{1,0.10} = 2.71$.

TABLE 4
CORRELATION COEFFICIENTS ACROSS BIVARIATE PROBIT MODELS FOR DIFFERENT FLOW LEVELS: PRIVATE BOATERS

	5 kcfs	13 kcfs	22 kcfs
13 kcfs	0.23		
	(0.23)		
22 kcfs	0.27	0.87	
	(0.18)	(0.08)***	
40 kcfs	0.32	0.84	0.63
	(0.17)*	(0.16)***	(0.17)***

Note: Numbers in ( ) indicate asymptotic standard errors.
*, **, *** indicate significance levels of 0.10, 0.05, and 0.01, respectively.

the joint model without restrictions. Indeed, the asymptotic standard errors on some of the coefficients actually increase with joint estimation, and the significance of the coefficients on Education (baseline permit), and the Miles 1 (15 percent fewer) crossed standard significance level thresholds. In contrast, an efficiency gain is observed for all the coefficients on the restricted variables as a result of imposing the cross-equation restrictions. In particular, the standard error on the bid variable falls noticeably.

Independent and joint estimation results for the "either sex" and the "antlerless" Sandhill hunting permits are provided in Table 3. The probability of a "yes" response increases with the perceived quality of the hunting experience, but declines with higher dichotomous choice posted prices. The coefficient on the year variable was only significant for the antlerless model, indicating that the 1992 respondents had higher values

TABLE 5
INDIVIDUAL AND JOINTLY ESTIMATED MEAN WTP DISTRIBUTIONS

	Distribution 1			Distribution 2			
ρ	Name	Mean (Indep)	Mean (Joint)	Name	Mean (Indep)	Mean (Joint)	$\sigma^2_{1-2,\text{JOINT}} / \sigma^2_{1-2,\text{INDEP}}$
0.32	40 kcfs	432	431	5 kcfs	243	237	0.84
		[369, 505]	[372, 493]		[198, 296]	[185, 298]	
0.39	Sandhill$_{ES}$	40.90	40.85	Sandhill$_A$	18.08	17.72	0.80
		[36.88, 45.51]	[36.64, 45.15]		[15.75, 21.86]	[15.43, 21.08]	
0.63	22 kcfs	522	527	40 kcfs	432	420	0.74
		[454, 599]	[454, 596]		[370, 504]	[352, 491]	
0.84	13 kcfs	518	528	40 kcfs	432	432	0.36
		[458, 588]	[466, 592]		[370, 504]	[377, 490]	
0.92	Escanaba$_{\text{Baseline}}$	5.40	5.51	Escanaba$_{\text{15\% fewer}}$	4.74	4.75	0.52
		[4.82, 6.08]	[4.93, 6.12]		[4.23, 5.32]	[4.22, 5.32]	

Note: Numbers in [ ] reflect the 0.90 confidence interval.

TABLE 6
SIGNIFICANCE LEVELS OF DIFFERENCE OF MEAN WTP ESTIMATES: SELECTED OBSERVATIONS

ρ	Comparison	$\hat{\alpha}_{indep}$	$\hat{\alpha}_{joint}$
0.63	22 kcfs–40 kcfs	0.066	0.026
0.84	13 kcfs–40 kcfs	0.069	0.003
0.92	Escanaba	0.093	0.016

for the antlerless permits. This result is consistent with the observation that the 1992 Sandhill hunt was limited to antlerless deer, and that the respondents generally reported a positive experience in spite of the fact that an antlerless hunt is popularly regarded to be inferior to an either sex hunt. Although the estimated correlation coefficient of 0.39 is much lower than the Escanaba study, it is still highly significant, indicating that any negative effects on correlation, if they exist, do not offset factors favoring a positive correlation. This lower correlation is reflected by the observation that there is an obvious difference in parameter estimates across scenarios. Notably, the effect of prices on WTP is more distinct for antlerless permits, suggesting both a lower value and variance in values for WTP in the antlerless scenario.

All possible combinations of individual and joint coefficient restrictions across the Sandhill equations were rejected using likelihood ratio tests with the unrestricted model as a reference. This demonstrates that entire valuation functions can be significantly different even though a significant correlation across equations is observed. Like the Escanaba study, the asymptotic standard errors of the joint-unrestricted model are quite similar to those of the independent model —indicating little efficiency gains from joint estimation without cross-equation equality restrictions.

The individual, the joint-unrestricted, and the joint-restricted models were estimated for each of the six possible *Grand Canyon White Water Boating Survey* valuation comparisons (5 vs 13 kcfs, 5 vs 22 kcfs, 5 vs 40 kcfs, 13 vs 22 kcfs, 13 vs 40 kcfs, and 22 vs 40 kcfs).[7] In two comparisons (5 vs 13 kcfs and 5 vs 22 kcfs) the hypothesis of no correlation across equations could not be rejected, and the two valuation response functions are statistically independent. Comparison of flow descriptions suggests that this result is not surprising, as the description of the 5 kcfs low flow scenario differs considerably from those at the more desirable moderate levels. As noted previously, some of the negative attributes of the 5 and the 40 kcfs scenarios were similar, which is consistent with the result that the correlation coefficient between WTP functions for these two scenarios was significant at the 10 percent level. All other pairwise correlations were significant at the 1 percent level.

Rejection of cross-equation equality restrictions also varied across the four pairwise comparisons that have significant correlation coefficients. The joint hypotheses that all coefficients were equal could not be rejected at the 1 percent level for the 13 vs 22 kcfs comparison, indicating that these flows have statistically similar valuation functions when correlation is accounted for. Hypotheses of equality across equations of the bid coefficient, the bid and constant coefficient, and the bid and cost coefficients could not be rejected for the 5 vs 40, the 13 vs 40, and the 22 vs 40 kcfs pairwise comparisons, respectively. Consistent with previous findings, some efficiency gains in terms of the individual coefficients were found when moving from the independent to the joint models.

Taken together, the results presented so far demonstrate that there can be significant correlation between responses to contingent

[7] As noted by an anonymous reviewer and W-133 participants, a more complete model would estimate all four scenarios simultaneously. Instead, bivariate probit models were estimated for each of the pairwise comparisons in this analysis. In addition, some caution should be taken in interpreting the magnitude of the correlation found in this analysis because the original study involved possible effects associated with different orderings of the contingent valuation questions (see Boyle, Welsh, and Bishop 1991, 1993). High correlations between scenarios may be attributed to proximity in the survey rather than to similarity in conditions. Similarly, low correlation values might be attributed to ordering effects. In spite of these limitations, the paired comparisons and the data are retained in this manuscript for illustrative purposes.

valuation questions elicited in the same survey. There are some indicators that this correlation is quite high when the attributes of the commodity being valued are quite similar across questions and declines with dissimilarities. In spite of the fact that the attributes vary widely across scenarios, the sign of the correlation coefficient is either positive and significant, or not significantly different from zero. Given that some scenarios encompass very different attributes, the lack of any negative correlation coefficients suggests that systematic and positive correlation effects are dominant.[8] These results also demonstrate the cross-equation equality restrictions do reduce the standard error of the estimated coefficients, a finding that is consistent with previous research (e.g., Alberini and Kanninen 1994).

Table 5 provides summary statistics for the estimated distributions of mean WTP for the individual and joint models for cases where the correlation coefficient was significant at the 10 percent level or better, and joint equality restrictions across equations for all elements of the coefficient vectors could not be rejected. The first column provides the correlation coefficient. The next three columns identify the distributions for the first scenario being compared and provide the bootstrap results from the joint and independent estimations of mean WTP using the "best" joint-restricted model in which cross-equation equality restrictions cannot be rejected. The fifth through seventh columns present the same information for the second scenario. The final column provides a ratio of the variance of the distribution of the differences from the joint model ($\sigma^2_{1-2,\ \mathrm{JOINT}}$) to the variance of the difference from the independent model ($\sigma^2_{1-2,\ \mathrm{INDEP}}$). This relationship is of particular interest because positive correlation in estimated mean WTP values is expected to reduce the variance of the difference, as suggested in the introduction.

A comparison across columns in Table 5 shows that even when the correlation coefficient is relatively high and cross-equation equality restrictions improve the efficiency of individual coefficient estimates, the estimation of joint models has only a very small effect on distributions of estimated mean WTP. Confidence ranges change only slightly, if at all, between the independent and joint models, suggesting that efficiency effects on the variance of the difference are minor.[9] However, the ratio of the joint to independent variance of the difference does decline with increases in the level of correlation. Combined, these results suggest that as correlations rise, the variance of the difference in estimated mean WTP distributions will decline due to correlation effects even though there may be negligible efficiency gains in terms of estimating individual mean WTP distributions.

Table 6 indicates that accounting for the correlation in the estimation process does impact on difference of means tests for comparisons in which the significance of the difference fell in statistically interesting ranges (i.e., around 0.10, 0.05, 0.01). Values associated with other comparisons in Table 5 diverged substantially from these critical values (e.g., 0.45 or 0.000000001) and, thus, are not that interesting in terms of this analysis. The values provided in Table 6 demonstrate that accounting for the correlation does have an effect on the decision of whether to accept or reject the hypothesis that the mean WTP is significantly different across scenarios. In each case presented it causes the estimated significance values to cross the 5 percent level, and, in the 13 kcfs−40 kcfs comparison, the critical significance level changes from 0.1 to 0.01. As such, joint estimation appears to have potentially important consequences from a policy perspective.

[8] That substitutability and negative correlation across question responses can occur in a random utility bivariate probit framework is demonstrated in a study by Horowitz (1994) of consumer preferences for government programs that reduce anthropogenic risks. In that study, however, substitutability was forced by explicitly altering the relative risks of competing programs and relegating the number of lives saved assumed by respondents (which are not observable) into the error term.

[9] Across the ten possible comparisons, the independent to joint ratio of the variance of mean WTP distributions averaged 0.9372 (0.1509) but was not significantly different from unity.

V. IMPLICATIONS AND CONCLUSIONS

The high correlation coefficients observed and the consequent effects on the difference in estimated mean WTP distributions indicate that a cross-scenario correlation may be an important factor in some policy comparisons, and is particularly high in closely related and embedded scenarios. As such, this paper provides empirical evidence from actual CV studies that standard assumptions of independence in comparing distributions lead to biased estimates of the difference, and may lead to erroneous conclusions about the significance of the difference in mean WTP values elicited in the same questionnaire.

From an applied perspective this empirical result must be weighed against the additional programming costs of implementing the joint estimator. If the cost of adopting the more complex bivariate probit approach is perceived to be high for the individual researcher, it is critical to recognize that there are instances in which the additional programming costs might not be warranted from a difference of means perspective. For example, a rule of thumb might be to not use joint estimation if independent mean WTP distributions overlap considerably, say at the 20 percent level or higher. Under these conditions it is unlikely that joint estimation will change the decision to not reject the null hypothesis of equality. At the other extreme, distributions that do not overlap at all when estimated independently, would indicate that—unless there was a strong reason to believe that responses are highly negatively correlated—the bivariate probit estimation approach would not change the hypothesis test results. At the same time, it should be acknowledged that the additional costs of joint estimation should not be prohibitive. Standard statistical packages such as LIMDEP have readily accessible bivariate routines.

The effort required to impose cross-equation restrictions may be much larger. Cross-equation equality restrictions are not an option in most standard statistical packages, and thus the researcher must possess more sophisticated programming skills in order to impose these restrictions.[10] Furthermore, the gains in precision from imposing cross-equation equality restrictions may be small. The results from this paper support previous research that such restrictions do increase the efficiency of estimated coefficients, but suggest that the indirect effects on estimated mean WTP distributions and their difference are modest at best. These mixed results may be attributed, in part, to the fact that the restricted coefficients were already highly significant in the individually estimated models, thus moderating the impact of additional efficiency gains. Based on the data used in this research, we cannot conclude that joint estimation will provide modest gains in efficiency of estimated distributions of mean WTP in all cases. In instances where the individually estimated models are not highly significant, joint estimation may provide larger efficiency gains than realized here.

In all, more empirical research is warranted before conclusions could be drawn about the importance of efficiency effects. Similarly, the source of correlation is not isolated in this study, and future research should also be directed towards identifying whether correlation is attributable to perceived similarities in attributes across scenarios or to a number of psychological response factors that have recently been suggested in the literature.

APPENDIX

Text of Escanaba Lake Survey "Full Permit"

28. Suppose that a special permit will be required this year to fish Escanaba Lake *during the "early season"*. Assume that you can order a permit to fish the "early season" at Escanaba Lake by mail. A permit is valid for the two weeks before the open of the regular fishing season (April 17 through May 1, 1992).

[10] Programming in this paper was conducted in Gauss 3.11, using the maxlik application.

- All individuals 16 years old and older wishing to fish Escanaba Lake between April 17 and May 1, 1992 will have to show a permit at the research contact station.
- Individuals with a permit may fish Escanaba Lake as often as they would like between April 17 and May 1, 1992.
- All the regulations currently in force for fishing Escanaba Lake will stay the same as they are now.
- If the ice goes out early, fishing will be free until April 17.
- The revenue from permit sales will go to the Northern Highland Fishery Research Area.

The amount we ask about below may seem very high or low to you *but it's very important that you answer the question seriously*. The amount written below was randomly assigned to you.

Would you pay $__ for a permit to fish Escanaba Lake between April 17 and May 1, 1992? (CIRCLE ONE NUMBER)

1. No
2. Yes

29. Assume that permits would be sold as described in question 28. In addition, assume that there will be about 15 percent fewer walleye in Escanaba Lake than usual at the beginning of the "early season". The size of the fish would be the same as now. It is hard to say how this would affect the catch of any one angler. Some anglers may catch just as many fish as usual while others may not do as well. In thinking about how this might affect your success, assume that there will be somewhat fewer walleyes around. As in question 28, the amount written below was randomly assigned to you.

Under these new conditions, would you pay $__ for a permit to fish Escanaba Lake between April 17 and May 1, 1992? (CIRCLE ONE NUMBER)

1. No
2. Yes

Text of Sandhill Public Deer Hunt Survey (1991)

16. Suppose that next year you apply for a Sandhill General Public Deer Hunt permit but are not chosen to receive a permit. Imagine that as part of a research project you have the chance to *purchase* an either sex permit. If you were able to buy a 1992 Sandhill *either sex* permit, would you be willing to pay $__? (CIRCLE ONE NUMBER)

1. No
2. Yes

17. If you were able to buy a 1992 Sandhill *antlerless* permit, would you be willing to pay $__? (CIRCLE ONE NUMBER)

1. No
2. Yes

Text of Grand Canyon White Water Boater Survey (22 kcfs)

At moderately high water levels (around 22,000 cfs), the pace of the river is faster than at lower flows, leaving more time for side canyons and stops at attractions. Boating groups do not have a problem staying on schedule. Rapids have larger waves and provide a bigger "roller coaster" ride than at moderate water. Only a few passengers choose to walk around some of the bigger rapids for their safety. Some potential campsites are under water in some areas of the canyon, but generally campsites are plentiful although a bit smaller in size.

We would now like you to imagine that you are presently deciding whether or not to go on a Grand Canyon white water trip. Imagine that the trip would be the same as your last trip (e.g., the people, food, etc.) with *two exceptions*:

> The water level would be constant at 22,000 cfs (see description for Case 4 above)

AND

> *Your individual* costs for the trip increased by $__ (over the total cost you calculated on page 8, question A26)

D2. Would you go on this trip? (CIRCLE ONE NUMBER)

1. YES, I WOULD PAY THIS AMOUNT TO TAKE THE TRIP
2. NO, I WOULD NOT PAY THIS AMOUNT TO TAKE THE TRIP

References

Alberini, Anna. 1995. "Efficiency vs. Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models." *Journal of Environmental Economics and Management* 29 (2):169–80.

Alberini, Anna, and Barbara J. Kanninen. 1994. "Efficiency Gains from Joint Estimation: When Does a Second Equation Improve Estimation of the First?" Paper presented at the AERE/AAEA annual meetings, San Diego.

Balistreri, Edward, Gary McClelland, Gregory L. Poe, and William D. Schulze. 1996. "Can Hypothetical Questions Reveal True Values? A Laboratory Comparison of Dichotomous Choice and Open-Ended Contingent Values with Auction Values." Cornell Working Paper Series in Environmental and Resource Economics, 96–01, Cornell University.

Boyle, Kevin J., Michael P. Welsh, and Richard C. Bishop. 1991. "An Empirical Investigation of the Role of Question Order, Symbolic Effects, and Respondent Experience in Contingent-Valuation Studies." Department of Agricultural and Resource Economics, University of Maine.

———. 1993. "The Role of Question Order and Respondent Experience in Contingent-Valuation Studies." *Journal of Environmental Economics and Management* 25 (1):S80–S99.

Cameron, Trudy A., and Michelle D. James. 1987. "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys." *Review of Economics and Statistics* 69 (2):269–76.

Cameron, Trudy A., and John Quiggin. 1994. "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-Up' Questionnaire." *Journal of Environmental Economics and Management* 27 (3):218–34.

Carson, Richard T., and Robert C. Mitchell. 1995. "Sequencing and Nesting in Contingent Valuation Surveys." *Journal of Environmental Economics and Management* 28 (2):155–73.

Champ, Patricia A., and Richard C. Bishop. 1996. "Evidence on the Accuracy of Expenditures Reported in Recreational Surveys." *Journal of Agricultural and Resource Economics* 21 (2):150–59.

Desvousges, William F., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and K. Nichole Wilson. 1992. *Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy.* Research Triangle Park, NC: Research Triangle Institute.

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

Fahrmeir, Ludwig, and Gerhard Tutz. 1994. *Multivariate Statistical Modelling Based on Generalized Linear Modelling.* New York: Springer-Verlag.

Goldberger, Arthur S. 1991. *A Course in Econometrics.* Cambridge: Harvard University Press.

Greene, William H. 1993. *Econometric Analysis.* 2d ed. New York: Macmillan Publishing Co.

Hanemann, W. Michael. 1984. "Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66 (3):332–41.

———. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses: Reply." *American Journal of Agricultural Economics* 71 (4):1057–61.

Hoehn, John P., and John B. Loomis. 1993. "Substitution Effects in the Valuation of Multiple Environmental Programs." *Journal of Environmental Economics and Management* 25 (1):56–75.

Horowitz, John. 1994. "Preferences for Pesticide Regulation." *American Journal of Agricultural Economics* 76 (3):394–406.

Hsiao, Cheng. 1986. *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Kahneman, Daniel, and Jack L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22 (1):57–70.

Kanninen, Barbara J. 1995. "Bias in Discrete Response Contingent Valuation." *Journal of Environmental Economics and Management* 28 (1):114–25.

Krinsky, Itzhak, and A. Leslie Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics* 68 (4):715–19.

Park, Timothy, John B. Loomis, and Michael Creel. 1991. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies." *Land Economics* 67 (1):64–73.

Poe, Gregory L., Eric K. Severance-Lossin, and Michael P. Welsh. 1994. "Measuring the Difference $(X - Y)$ of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76 (4):904–15.