TESTING FOR HYPOTHETICAL BIAS IN CONTINGENT VALUATION

USING A LATENT CHOICE MULTINOMIAL LOGIT MODEL

by

Steven B. Caudill[1]

Peter A. Groothuis[2]

And

John C. Whitehead[3]

April 27, 2006

[1]Department of Economics, Auburn University, AL 36949, scaudill@business.auburn.edu

[2]Department of Economics, Appalachian State University, groothuispa@appstate.edu

[3] Department of Economics, Appalachian State University, whiteheadjc@appstate.edu

Testing for Hypothetical Bias in Contingent Valuation
Using A Latent Choice Multinomial Logit Model

Abstract

The most persistently troubling empirical result in the contingent valuation method literature is the tendency for hypothetical willingness to pay to overestimate real willingness to pay. We suggest a new approach to test and correct for hypothetical bias using a latent choice multinomial logit (LCMNL) model. To develop this model, we extend Dempster, Laird, and Rubin's (1977) work on the EM algorithm to the estimation of a multinomial logit model with missing information on categorical membership. Using data on both the quality of water in the Catawba River in North Carolina and the preservation of Saginaw wetlands in Michigan, we find two types of "yes" responders in both data sets. We suggest that one set of yes responses are yea-sayers who suffer from hypothetical bias and answer yes to the hypothetical question but would not pay the bid amount if it were real. The second group does not suffer from hypothetical bias and would pay the bid amount if it were real.

Testing for Hypothetical Bias in Contingent Valuation:
Using A Latent Choice Multinomial Logit Model

Introduction

The most persistently troubling empirical result in the CVM literature is hypothetical

bias, the tendency for hypothetical willingness to pay to overestimate real willingness to pay

(Cummings, Harrison, and Rutström, 1995, Cummings, Elliot, Harrison, and Murphy 1997,

Blumenschein, Johannesson, Blomquist, Liljas and O'Conor 1997). Hypothetical bias occurs

when CVM respondents state that they will pay for a good when in fact they will not, or they will

actually pay less, when placed in a similar purchase decision. Hypothetical bias is usually

attributed to the presence of passive use values and lack of familiarity with paying for policies

that provide passive use value. Hypothetical bias, however, has been found in a variety of

applications, including private goods for which no passive use values should exist (List and

Gallett, 2001). Surprisingly, hypothetical bias is ignored in much of the CVM literature

(Harrison, forthcoming).

Hypothetical bias arises because answers to CVM willingness-to-pay questions have no

real consequences other than a weak connection to the influence of government policy.

Respondents who state that they would pay for the policy change are not required to actually

pay. Some respondents may state that they would pay for the policy when, in fact, they would

not if placed in the real situation. There are at least two possible explanations for this behavior.

The respondent may be trying to influence policy by signaling support (e.g., strategic bias, warm

glow) and/or the respondent may simply be pleasing the interviewer (e.g., yea saying).

Hypothetical bias leads to upwardly biased willingness to pay estimates. Willingness to pay

estimates from the CVM must be considered upper bounds of benefits in the context of benefit-cost analysis unless steps are taken to mitigate hypothetical bias. In essence, the existence of hypothetical bias means that in CVM there are two types of yes responders that need to be identified and separated.

To test for hypothetical bias, we develop a latent choice multinomial logit model to separate the yes responses into two categories. In an important paper, Dempster, Laird, and Rubin (henceforth DLR) (1977) show how the EM algorithm can be used to obtain maximum likelihood estimates from incomplete data. The first illustration discussed by DLR is a multinomial probability example using data from Rao (1955) on 197 animals divided into four categories. DLR consider the case where the original first category of animal is split into two new categories, but exactly which animals are assigned to which category is unknown. The resulting multinomial probability model is characterized by five categories and missing information on membership in categories one and two. DLR show how the parameters of this model can be estimated by maximum likelihood using the EM algorithm. This paper extends DLR's work on the use of the EM algorithm to the estimation of a multinomial logit model (henceforth MNL) with latent or hidden choices to the estimation of a latent choice multinomial logit model (henceforth LCMNL).[1] The usual multinomial logit model is a special case of the LCMNL. The LCMNL is related to the latent class multinomial logit model discussed by Greene and Hensher (2002) in which the model is composed of multinomial logit models which differ across individuals according to class membership, which is unknown. In the LCMNL there are no unknown classes but the choices of some individuals are not observed.

To test for hypothetical bias the LCMNL model is applied to two contingent valuation

data sets. The first is based on responses to questions about water quality in the Catawba River in North Carolina while the second is based on response to questions about wetland preservation in Saginaw watershed in Michigan. Application of the LCMNL model to the two WTP data reveals two types of "yes" responders in both data sets.

A Latent Choice Multinomial Logit Model

To discuss the LCMNL model and to facilitate comparisons with the MNL model, we adopt the language associated with the nested logit model. That is, we characterize the choice model in terms of branches and stems. Branches are observed alternatives and stems are unobserved alternatives associated with branches. A branch may contain any number of unobservable stems, including zero. Thus, the usual MNL logit model is a LCMNL having only observable branches, each with no stems.

As a point of departure, we begin with the usual MNL model with m observable choices or branches. Probabilities in this model are given by

$$P_{ij} = \exp[X_i \beta_j]/D \; where \; D = \sum_{j=1}^{m} \exp[X_i \beta_j]. \tag{1}$$

The data include a set of indicator or responses variables denoting the choice made by each individual. Let $Y_{ij}=1$ if individual i makes choice j and $Y_{ij}=0$, otherwise. This leads to the usual loglikelihood function and the familiar first order conditions

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} \log P_{ij} \tag{2}$$

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^{n} (Y_{ij} - P_{ij}) X_i', \; for \; j = 1,...,m \;. \tag{3}$$

In the LCMNL model, there is the possibility that some of the observed branches contain unobserved stems. To develop this model, we denote the unknown parameters by $\exists_{jk}$ where j indicates "branch" and k indicates "stem." Probabilities in this model are given by

$$P_{ijk} = \exp[X_i\beta_{jk}]/D \; where \; D = \sum_{j=1}^{m}\sum_{k=1}^{s_j}\exp[X_i\beta_{jk}], \qquad (4)$$

where m represents the number of branches in the model and s_j represents the number of stems associated with branch j. If $s_j = 1$ for all branches (no branch has any stems), the usual multinomial logit model obtains.[2]

The loglikelihood function for the LCMNL model is based on probabilities like those above in (4) and is given by

$$\log L = \sum_{i=1}^{n}\sum_{j=1}^{m}Y_{ij}\log\left[\sum_{k=1}^{s_j}P_{ijk}\right]. \qquad (5)$$

The indicator variable, Y_{ij}, indicates only which branch is chosen as stem membership is unknown. That is why the logarithm of the sum of the probabilities on a branch appears in the likelihood function. Once again, it is clear that if $s_j=1$ for all branches, the loglikelihood function for the usual multinomial logit model given in (2) obtains.

Maximization of the likelihood function for the LCMNL requires the first derivatives with respect to the unknown parameters in the model. The first derivatives can be shown to equal

$$\frac{\partial \log L}{\partial \beta_{jk}} = \sum_{i=1}^{n}\alpha_{ijk}(Y_{ij}-P_{ij})X_i', \; for \; j=1,...,m \quad k=1,...,s_j, \qquad (6)$$

where the construct, $\forall_{ijk}$, is defined

$$\alpha_{ijk} = \frac{P_{ijk}}{\sum\limits_{k=1}^{s_j} P_{ijk}} = \frac{P_{ijk}}{P_{ij}} \ letting \ \sum\limits_{k=1}^{s_k} P_{ijk} = P_{ij}. \tag{7}$$

The construct, $\forall_{ijk}$, is the conditional probability of stem k given branch j calculated for each individual. Note that if the $\forall$s equal one for all individuals and all stems, the first order conditions for the usual multinomial logit model obtain.

A look at the second order conditions of the LCMNL model is also instructive. Three second derivatives are required to characterize the LCMNL model. First, we need the second derivatives like

$$\frac{\partial^2 \log L}{\partial \beta_{jk}^2} = \sum_{i=1}^{n} [-\alpha_{ijk}^2 P_{ij}(1-P_{ij}) + \alpha_{ijk}(1-\alpha_{ijk})(Y_{ij}-P_{ij})]X_i X_i'. \tag{8}$$

The cross-partial within a branch is given by

$$\frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{jl}} = \sum_{i=1}^{n} -\alpha_{ijk}\alpha_{ijl}[P_{ij}(1-P_{ij}) + (Y_{ij}-P_{ij})]X_i X_i'. \tag{9}$$

The cross-partial across branches is given by

$$\frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{ml}} = \sum_{i=1}^{n} [\alpha_{ijk}\alpha_{iml} P_{ij} P_{im}]X_i X_i'. \tag{10}$$

Again notice that if $\forall_{ijk}=1$ when j=k and 0 otherwise, that is all choices are observed (no branch has any stems) the derivatives in (9), (10), and (11) become

$$\frac{\partial^2 \log L}{\partial \beta_{jk}^2} = \sum_{i=1}^{n} [-P_{ij}(1-P_{ij})]X_i X_i' \tag{11}$$

$$\frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{jl}} = 0 \tag{12}$$

$$\frac{\partial^2 \log L}{\partial \beta_{jk} \partial \beta_{ml}} = \sum_{i=1}^{n} [P_{ij} P_{im}] X_i X_i^{'}. \tag{13}$$

Again, these are exactly the second derivatives of the usual multinomial logit model.

In the absence of any restrictions on the model there is an identification problem. This problem

occurs because the value of the likelihood function in (5) is unaffected by any reordering of

probabilities associated with stems on a particular branch. This problem can be solved by

imposing the following constraint, applied to the stems on each branch j

$$\sum_i \alpha_{j1} < \sum_i \alpha_{j2} < ... < \sum_i \alpha_{js_j}. \tag{14}$$

This constraint has the effect of ordering stems on each branch from least likely to most likely.

This constraint can be imposed after estimation by merely reordering the stems on a given

branch. This is similar to the approach used to identify the components of a mixture by Aitken

and Rubin (1985).

A Specific Case. In this section the LCMNL model is adapted to estimate a model with

two branches, with one branch containing two stems. This model is applied to contingent

valuation WTP data. We begin with a MNL model in which individuals make one of three

choices. Probabilities in this model are given by

$$P_{ij} = \frac{\exp(X_i \beta_j)}{\sum_{j=1}^{3} \exp(X_i \beta_j)} \quad j = 1,...,3, \tag{15}$$

where the β_j's are parameters to be estimated and X_i is a vector of exogenous variables. The

usual normalization applies so that $\exists_3 = 0$. The alternative selected in this model is indicated with

the usual set of dummy variables: Y_{i1}, Y_{i2}, and Y_{i3}, each taking the value zero or one, indicating that an alternative was or was not selected. The loglikelihood function in this MNL model is given by

$$\log L = \sum_{i=1}^{n} (Y_{i1} \log(P_{i1}) + Y_{i2} \log(P_{i2}) + Y_{i3} \log(P_{i3})), \tag{16}$$

where n is the sample size. Maximum likelihood estimation of the parameters in this model is routine. The first order conditions for maximization are

$$\frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^{n} (Y_{ij} - P_{ij}) X_i' = 0 \quad j = 1, 2. \tag{17}$$

We now assume the model has hidden choices. In particular, we assume the model has two branches and one branch has two stems. We denote the probabilities associated with the first branch by P_1 and probabilities on the stems associated with the second branch by P_{21} and P_{22}. The probability definitions are essentially the same as with the MNL model with three observed choices but the likelihood function has changed. This change occurs because only the branch choice, Y_2, is observed. The resulting incomplete-data/observed loglikelihood function is a special case of that given in (5)

$$\log L = \sum_{i=1}^{n} [Y_{i1} \log(P_{i1}) + Y_{i2} \log(P_{i21} + P_{i22})]. \tag{18}$$

The first order condition for the parameters associated with the first branch are the same as in the MNL model,

$$\frac{\partial \log L}{\partial \beta_1} = \sum_{i=1}^{n} (Y_{i1} - P_{i1}) X_i' = 0. \tag{19}$$

The first order condition for the parameters associated with the first stem on branch two is given

by

$$\frac{\partial \log L}{\partial \beta_{21}} = \sum_{i=1}^{n} \alpha_{i21}(Y_{ij} - P_{ij})X_i^{'} = 0,$$

$$where\ \alpha_{i21} = \frac{P_{i21}}{P_{i21} + P_{i22}}, i = 1,...,n.$$

(20)

The first order condition (if it were needed) for the parameters of the second stem associated

with branch 2 is similarly defined.

Estimation of the model by maximum likelihood is relatively straightforward if the

expectations maximization, or EM algorithm, is used. In the E step of the EM algorithm the

latent or missing variables are replaced by their conditional expectations given the data and

initial parameter estimates. The likelihood function is then maximized to obtain new parameter

values. These values can be used to obtain new conditional expectations and the process is

repeated.

To use the EM algorithm to estimate the simple LCMNL model, we denote the set of

unobservable indicator variables associated with the two stems on branch two by Y_{i21}^{*} and Y_{i22}^{*}.

These dummy variables take the value one if the observation is associated with that stem and

zero otherwise. Using these unobservable variables, the complete data loglikelihood in our

simple problem can be written

$$\log L = \sum_{i=1}^{n} (Y_{i1} \log(P_{i1}) + Y_{i21}^{*} \log(P_{i21}) + Y_{i22}^{*} \log(P_{i22})).$$

(21)

The likelihood function above characterizes the estimation problem as a missing data problem.

If the unobserved Y^{*}s were known, estimation would be as simple as estimating a MNL model.

In the expectations or "E" step of the EM algorithm, the unobserved Y^{*}s are replaced with their

conditional expectations given the data and values of the unknown parameters. The conditional

expectation is the probability of a stem, given the branch. These conditional expectations or

probabilities are well-known in the logit model and are given by

$$E(Y_{i21}^{*} \mid Y_{i2} = 1) = \frac{\exp(X_i\beta_{21})}{\exp(X_i\beta_{21}) + \exp(X_i\beta_{22})}$$

$$E(Y_{i22}^{*} \mid Y_{i2} = 1) = \frac{\exp(X_i\beta_{22})}{\exp(X_i\beta_{21}) + \exp(X_i\beta_{22})}$$

(22)

With the conditional expectations inserted into the loglikelihood function, the maximization or

"M" step of the EM algorithm maximizes the loglikelihood function. New parameter values are

obtained, then the "E" step and the "M" step are repeated. This process continues until the

likelihood function is maximized. Once the maximum has been found, standard errors are

calculated using one iteration of the algorithm of Berndt, Hall, Hall, and Hausman (1974).

One can see that the EM algorithm is embedded in the first order conditions given in (6)

and (20) above. The αs represent, for each individual, the conditional expectation of each stem,

given the branch. Thus, the αs are the conditional expectations given above in (22). The EM

algorithm uses initial parameter estimates to calculate the αs in (6) or (20) and then maximizes

the likelihood function. New parameter values are generated, new αs, and the process is

repeated.

The ability to locate hidden stems in a multinomial logit model is a useful result that can

lead to the estimation of logit models under many different scenarios. In the unconstrained form,

just presented, the LCMNL model describes hidden choices or individual heterogeneity. By

using the constraint for pooling choices in a MNL model given by Cramer and Ridder (1991),

many other possibilities occur. Suppose, for example, that the reason that a branch has two

stems is because one of the stems represents responses that have been misclassified from another branch. Imposing the constraint of Cramer and Ridder in the estimation leads to reclassification of the response data. Used this way, the LCMNL model becomes a reparameterization of the logit model for misclassification given by Hausman, Abrevaya, and Scott-Morton (1998). As a model of misclassification, the LCMNL model has been applied to the problem of fraud detection by Caudill, Ayuso, and Guillen (2005), hidden unemployment by Caudill (2003), and the problem of misclassified CV responses by Caudill and Groothuis (2005).

With different parameter constraints imposed, other response configurations are possible. Suppose a model has two observable branches, each having two unobservable stems. Suppose that one of the stems from each branch is associated with a third, unobservable branch. If the constraint of Cramer and Ridder is used, a logit model with three alternatives can be estimated from data on two observable choices. This paper is the first application of the unconstrained version of the LCMNL. Our application examines whether there are two categories of yes respondents in a contingent valuation survey that arises because of hypothetical bias.

Application to CV Data

In the past, two approaches have been developed to mitigate the overstatement of hypothetical willingness to pay. The ex-ante approach addresses hypothetical bias in the survey design stage. Respondents are variously (a) told that there are substitutes for the policy available, (b) reminded that they are income constrained, (c) asked to answer as if they were placed in an actual payment situation, and (d) told that hypothetical bias is a significant problem and asked not to succumb to this type of respondent error. The ex-post approach addresses hypothetical bias with follow-up questions to the hypothetical willingness to pay question.

Respondents who indicate that they are willing to pay for the policy are asked to rate the certainty they have in their willingness to pay. Respondent certainty is measured on a qualitative or quantitative scale where the low and high ends of the scale allow respondents to express their degree of certainty about their payment. Hypothetical willingness to pay responses are then recoded based on the certainty of the respondent.

Ex-Ante Approach

The ex-ante approach to hypothetical bias mitigation has evolved from simple reminders to respondents about economic constraints to elaborate lessons on how to avoid overstating one's willingness to pay. Loomis, Gonzalez-Caban, and Gregory (1994) reminded respondents about substitutes and income constraints and find that these reminders do not affect willingness to pay. In Loomis, Brown, Lucero, and Peterson (1996) respondents are reminded about income constraints and asked to answer as if they would actually pay. The authors find that the additional survey information moves hypothetical willingness to pay towards real willingness to pay.

Cummings and Taylor (1999), in what is called a "cheap talk" script, define hypothetical bias for respondents, explain why it may occur, and ask respondents to behave as if they are in a real payment situation. The authors find that the divergence between hypothetical and real willingness pay is eliminated by the cheap talk script. List and Gallet (2001) finds that the cheap talk script eliminates hypothetical bias for utility maximizers (card show consumers) but not for profit maximizers (card show dealers) who have more familiarity with the value of the product. Brown, Ajzen, and Hrubes (2003) find that a cheap talk script is able to mitigate hypothetical

bias at high bid levels but not at low bid levels. Aadland and Caplan (2003) find that a "short-scripted" cheap talk design for phone surveys is able to mitigate hypothetical bias for respondents who have strong environmental preferences. Lusk (2003) finds that the cheap talk script eliminates hypothetical bias for respondents with less knowledge about the goods. The problem with this approach is there usually no way to detect if the reminder was effective. Our approach can be used to see if the ex-ante approach eliminates hypothetical bias by testing to see if two groups of yes respondents exist.

Ex-Post Approach

Studies that have employed the ex-post correction approach have used qualitative and quantitative certainty scales. Johannesson, Liljas, and Johansson (1998) use two certainty categories and consider only those respondents who indicate they are "absolutely sure" about payment (the other category is "fairly sure"). They find that hypothetical willingness to pay understates real willingness to pay when only those who are "absolutely sure" are considered. Blumenschein, Johannesson, Blomquist, Liljas, O'Conor (1998), Blumenschein, Johannesson, Yokoyama, and Freeman (2002), and Blumenschein, Blomquist, Johannesson, Horn, and Freeman (2004) consider only those respondents who indicate they are "definitely sure" about payment (the other category is "probably sure"). These studies find that hypothetical willingness to pay is no different than actual willingness to pay when adjusted by respondent certainty.

A number of studies have used quantitative certainty scales. Champ, Bishop, Brown, and McCollum (1997) consider only those respondents who indicate that they are very certain at the highest point of a ten point quantitative scale (i.e., ten is very certain). The percentages of

respondents who are very certain about paying and respondents who would actually pay are no different. Champ and Bishop (2001) find that those respondents who are certain of their willingness to pay at the 8 or higher level on a ten point scale have a hypothetical willingness to pay similar to a real willingness to pay sample. Poe, Clark, Rondeau, and Schulze (2002) and Vossler, Ethier, Poe, and Welsh (2003) find that those respondents who rate their certainty of their willingness to pay at seven or higher on a ten point scale have probabilities of payment similar what is typically found in a real willingness to pay sample. Johannesson, Blomquist, Blumenschein, Johansson, Liljas, and O'Conor (1999) find respondents who are willing to pay in the hypothetical treatment are more likely to actually pay as their certainty rating rises.

A few studies have considered both ex-ante and ex-post approaches. Poe, Clark, Rondeau, and Schulze (2002) include a short cheap talk script and find that it has no effect on willingness to pay. Aadland and Caplan (2003) include a three-level qualitative certainty rating question but do not recode yes responses. Blumenschein, Blomquist, Johannesson, Horn, and Freeman (2004) find that cheap talk does not mitigate hypothetical bias but the use of certainty ratings eliminates the bias. Although both the ex-ante and ex-post approaches have been somewhat successful in the past, each has shortcomings. The ex-ante approach does not always work and there is no way to identify when it will work and when it will not. The ex-post approach occasionally works, but suffers from ad-hoc cutoff assignments. Our approach complements the ex-ante and ex-post approaches by providing a test to see if two categories of yes responses exist in the data. If two categories are found, we can use the above approaches as indicators on whether hypothetical bias exits.

We use two data sets to illustrate an application of the LCMNL technique to test for hypothetical-bias in CV data. The first data set used in this study is contingent valuation data on water quality in the Catawba watershed in North and South Carolina. The data has been previously examined by Eisen-Hecht and Kramer (2002) and Kramer and Eisen-Hecht (2002). The second data set used is contingent valuation data on the preservation of wetlands in the Saginaw watershed in Michigan. In both data sets we find that the LCMNL approach separates the yes responses into two categories. We report the results of the Catawba watershed first.

Catawba Watershed

The CV question for the Catawba study involved a management plan for protecting water quality in the Catawba basin. The management plan was offered at eight different price levels from $5 to $250 and respondents were asked about their willingness-to-pay. The data set contains 915 responses. Of these responses, 68 percent indicated a willingness to pay the specified amount (32 percent did not). A description of the variables used, along with summary statistics is given in Table 1. TAXYN is a dummy variable equal to one if the respondent rated reducing state and federal taxes as important. WPCONTROL is a dummy variable equal to one if the respondent had previously heard of efforts to control water pollution. USEYN is a dummy variable equal to one if the individuals rated their own use of the Catawba River as an important reason why the management plan would be of value to them. DRQUALYN is a dummy variable equal to one if the respondent rated the quality of their drinking water an important reason why the management plan would be of value to them. OTHUSEYN is a dummy variable equal to one if the respondent rated the use of the Catawba River by their friends and family as an important reason why the management plan would be of value to them. EXISTYN is a dummy variable

equal to one if the respondent rated the knowledge that the water quality in the basin was being protected regardless of their use of it as an important reason why the plan would be of value to them. LIKELYYN is a dummy variable equal to one if the respondent thought the management plan was somewhat or very likely to succeed. ITEMYN is a dummy variable equal to one if the respondent owns at least one item for outdoor water-based recreation. ENVORG is a dummy variable equal to one if the respondent belonged to an environmental or conservation organization. QUALWORS is a dummy variable equal to one if the respondent thought the water quality in their area had gotten worse over the last five years. TAPGOOD is a dummy variable equal to one if the respondent thought their tap water was above average or excellent in quality. AGE is the age of the respondent. DATELAG is the number of days between when the information booklet was mailed to the respondent when the interview was conducted. NEWAREA is a dummy variable equal to one if the respondent lived in the basin less than five years. UNIVYN is a dummy variable equal to one if the respondent completely trusted universities. EDUYN is a dummy variable equal to one if the respondent had completed some college or higher. SEX is a dummy variable equal to one if the respondent is male. INCOME is the household income of the respondent.

As a first step in the analysis, a simple logit model of the WTP decision is estimated. In this analysis, the dependent variable is coded 1=NO. The estimation results are given in Table 2. The parameter estimates are given in column 2 of Table 2. Several of the coefficients are statistically significant. In particular, the coefficients of WTPAMT, TAXYN, OTHUSEYN, EXISTYN, LIKELYYN, ENVORG, TAPGOOD, NEWAREA, UNIVYN, EDUYN, and INCOME are all significant at the $\forall$=.10 level or better. The marginal effects are given in

column 3 of Table 2 and exhibit the same pattern of statistically significant results. As these results are estimated for the sake of comparison, our discussion is limited as we turn our attention to the LCMNL estimation results.

The LCMNL model estimated here has two branches, with one branch having two stems. In our first effort to estimate the model we allowed for two types of "No" response. In this data set we were not able to find two types of "No" responders. We next estimate the model allowing for the possibility of two types of "Yes" responders. The LCMNL model did indeed find two types of "Yes" responders. We arbitrarily designate these two types of responders Yes1 and Yes2. The probability of being a Yes1 responder is estimated to be 0.255 and the probability of being a Yes2 responder is estimated to be 0.423. The results of the estimation of this model are given in Table 3 with the coefficients of the Yes2 responders normalized to zero.

Column 2 of Table 3 presents the estimated coefficients for the "No" responders. The coefficients of WTPAMT, USEYN, LIKELYYN, ENVORG, QUALWORS, UNIVYN, EDUY, SEX, and INCOME are significantly different from zero at the $\forall$=.10 level or better. The results from estimating the parameters associated with the Yes1 responders are given in column 3 of Table 3. In the Yes1 equation, the coefficients of WTPAMT, USEYN, OTHUSEYN, EXISTYN, ITEMYN, ENVORG, QUALWORS, TAPGOOD, and SEX are significantly different from zero at the $\forall$=.10 level or better. We do not dwell on these results because the marginal effects have been calculated and are much easier to interpret.

The marginal effects associated with the LCMNL model are given in Table 4. Column 2 presents the marginal effects associated with the No responders. The marginal effects associated with WTPAMT, TAXYN, OTHUSEYN, EXISTYN, LIKELYYN, NEWAREA, UNIVYN,

EDUYN, and INCOME are all statistically significant at the $\forall=.10$ level or better. The pattern of statistical significance is very similar to the results for the binomial logit. What is also clear is from the comparison is that the coefficients of the LCMNL model are about twenty-five percent larger in absolute value than their binomial counterparts.

The marginal effects associated with the Yes1 responders are given in column 3 of Table 4. Marginal effects associated with WTPAMT, USEYN, OTHUSEYN, EXISTYN, ITEMYN, ENVORG, QUALWORS, TAPGOOD, and SEX are statistically significant at the $\forall=.10$ level or better. The marginal effects associated with the Yes2 reponders are givemn in column 4 of Table 4. Marginal effects associated with WTPAMT USEYN, OTHUSEYN, EXISTYN, LIKELYYN, ITEMYN, ENVORG, QUALWORS, TAPGOOD, EDUYN, and SEX are statistically significant at the $\forall=.10$ level or better. In many cases the directions of the marginal effects differ between the Yes1 responders and the Yes2 responders. There are several cases where the marginal effects are statistically significant at the $\forall=.10$ level or better but the signs differ. This happens for the coefficients of WTPAMT, USEYN, OTHUSEYN, EXISTYN, ITENYN, ENVORG, QUALWORS, TAPGOOD, and SEX.

In order to shed more light on differences between the Yes1 and Yes2 regimes, we examine the means of the independent variables for those individuals with the ten highest predicted probabilities from the LCMNL estimation associated with each "Yes" response. The means for these individuals are given in Table 5. In comparison to those individuals associated with the Yes2 regime had a much lower dollar amount of the management plan, were much more likely to rate their own use of the River as an important consideration, were far more likely to belong to an environmental organization, more likely to believe their water quality had gotten

worse, and earned about twice as much as those associated with Yes2. There appear to be two very different types of Yes responders to the CV question. The first set of responders that have characteristics that suggest that they would indeed be willing to pay the specified amount while the other category suggests that they suffer from hypothetical bias.

Saginaw Wetlands

The second data set used in this study is contingent valuation data on the preservation of wetlands in the Saginaw Bay watershed in Michigan. The data has been previously examined by Whitehead et al. (2005). The CV question involved a management plan for purchasing wetlands for preservation in the Saginaw Bay watershed. The purchase plan was offered at six different price levels from $25 to $200 and respondents were asked about their willingness-to-pay. The data set contains 281 responses. Of these responses, 55 percent indicated a willingness to pay the specified amount (45 percent did not). A description of the variables used, along with summary statistics is given in Table 6. LNBID is the natural log of the eight different price levels and ACRES is the number of acres of wetlands that are to be purchased to preserve. TRAVCOST is the cost to the respondent of travel to the Saginaw Bay area. SUBCOST is the cost to travel to a substitute site for recreation, either Traverse City on Lake Michigan or Alpena on Lake Huron, whichever was closer for the respondent. INCOME3 is the income of the respondent, while MEMBER is a dummy variable equal to one if the respondent was a member of an environmental group and LIKELY2 is a dummy variable equal to one if the respondent thought enough people would donate to preserve the Saginaw wetlands.

As before, we begin with the estimation of a simple logit model of the WTP decision. In this analysis, the dependent variable is coded 1=NO. The estimation results are given in Table 7. The parameter estimates are given in column 2 of Table 7. Several of the coefficients are statistically significant. In addition to the intercept, the coefficients of LNBID (+), INCOME3 (-), MEMBER (-), and LIKELY2 (-) are statistically significant at the $\forall=.10$ level or better and have the expected sign. The marginal effects associated with the explanatory variables are given in column 3 of Table 7 and exhibit the same pattern of statistically significance. As before, these results are estimated for the sake of comparison, so our discussion is limited as we turn our attention to the LCMNL estimation results.

We, again, estimate the model allowing for the possibility of two types of "Yes" responders and we, again, find empirical evidence for the presence of two types of "Yes" responders. The probability of being a Yes1 responder is estimated to be 0.342 and the probability of being a Yes2 responder is estimated to be 0.110. The results of the estimation of this model are given in Table 8 with the coefficients of the Yes2 responders normalized to zero.

Column 2 of Table 8 presents the estimated coefficients for the "No" responders. Only the coefficient of LNBID is significantly different from zero at the $\forall=.10$ level or better. The results from estimating the parameters associated with the Yes1 responders are given in column 3 of Table 3. In the Yes1 equation, none of the estimated coefficients is significantly different from zero at the $\forall=.10$ level or better.

The marginal effects associated with the LCMNL model are given in Table 9. Column 2 presents the marginal effects associated with the No responders. Only the marginal effects associated with LNBID and LIKELY2 are statistically significant at the $\forall=.10$ level or better.

The marginal effects associated with the Yes1 responders are given in column 3 of Table 9. Marginal effects associated with INCOME3 and MEMBER are statistically significant at the $\forall=.10$ level or better. The marginal effects associated with the Yes2 responders are given in column 4 of Table 9. None of the individual marginal effects is statistically significant at the $\forall=.10$ level or better. Although this finding is somewhat troubling, it may indicate the presence of people behaving "randomly" and not in accord with the usual economic theory of utility maximization.

As before, in order to shed more light on differences between the Yes1 and Yes2 regimes, we examine the means of the independent variables for those individuals with the ten highest predicted probabilities from the LCMNL estimation associated with each "Yes" response. The means for these individuals are given in Table 10. In comparison to the Yes2 responders, those individuals associated with the Yes1 regime had a much lower dollar amount of the bid, much higher income, have lower travel cost and are much more likely to belong to an environmental organization than those associated with Yes2. There again appear to be two very different types of Yes responders to the CV questioning this data set. To further explore the possibility of hypothetical bias, we focus on the intensity of preference follow-up question in the next section.

Hypothetical Bias Detection using the LCMNL model

In both the Catawba Watershed and the Saginaw Wetland CV surveys, two groups of yes respondents are found. When looking at characteristics of each set of yes respondents, one set of yes respondents seems to exhibit hypothetical bias. In this section, we focus on the Saginaw Wetland data to test whether the LCMNL procedure identifies the same respondents as indicated

by the intensity of preference correction. Intensity of preference correction data is not available in the Catawba data set. In the LCMNL model predicted probabilities are calculated for each individual responding No, Yes1, and Yes2. Using the follow-up questions in the Saginaw data set, we separate the responses into three categories: No responses, Yes1 responses, and Yes2 responses. Yes1 responses are individuals who answered yes to the CV question and seven or higher on a ten point scale to the question, "On a scale of 1 to 10, how sure are you that you would make the one-time donation?" Yes2 responses are individuals who answered yes to the CV question and less than seven on the certainty question. In order the test for convergent validity between the two hypothetical bias detection approaches, we perform two tests comparing the predicted probabilities from the LCMNL model to the probabilities determined from the follow-up question.

The first test compares the means of the predicted LCMNL probabilities to the response categories generated by the ex-post hypothetical bias identification. In Table 11a in columns two through four, we report the results of the comparison-of-means test. We find that for all three comparisons the LCMNL probabilities are significantly different and higher for respondents in the matched categories. In other words, respondents identified as yes1 respondents using the following up questions have higher LCMNL yes1 predicted probabilities than other respondents. The second test uses simple logit models with our constructed intensity of preference dummies as dependent variables and the predicted probabilities from the LCMNL model as the independent variable. These results are reported in Table 11b and also indicate the presence of positive and statistically significant associations between the predicted probabilities from the LCMNL model and ex-post classifications. Both tests presented here indicate that the LCMNL

probabilities align with the intensity of preference categories. In other words, respondents who have higher LCMNL yes1 probabilities are more likely to be categorized as yes1 respondents using the follow-up certainty question. These results indicate that the LCMNL model is indeed detecting respondents with hypothetical bias.

Conclusions

This paper uses the EM algorithm of Dempster, Laird, and Rubin (1977) to estimate a multinomial logit model with missing information, which we call the latent choice multinomial logit model, to test for hypothetical bias in CV analysis. The LCMNL model is applied to CV based on responses to questions about water quality in the Catawba River and the preservation of wetlands in the Saginaw Bay area. In both applications of the LCMNL model to the WTP data, we find two types of "yes" responders. Typically, one type is much wealthier and more environmentally conscious than the other. For the Saginaw data we show that one of the "yes" groups identified by the LCMNL model has much more conviction about saying yes that the other "yes" category. We suggest that this technique identifies responses associated with hypothetical bias and provides a mechanism for correcting this bias.

Footnotes

[1] A similar model has been estimated by Magder and Hughes (1997).

[2] The researcher must determine a priori the number and location of the unobservable stems in the model.

Table 1
Variable Definitions and Summary Statistics
Catawba data

VARIABLE NAMES AND DEFINITIONS	MEAN (ST DEV)
WTPAMT (dollar amount of management plan, from $5 to $250)	98.678 (86.07)
TAXYN (1 if respondent rated reducing state and federal taxes important to them, 0 otherwise	0.707 (0.46)
WPCONTRO (1 if respondent had previously heard of efforts to control water pollution, 0 otherwise)	0.715 (0.45)
USEYN (1 if respondent rated their own use of the Catawba River as an important reason why the management plan would be of value to them, 0 otherwise)	0.550 (0.50)
DRQUALYN (1 if respondent rated the quality of the drinking water in their area as an important reason why the management plan would be of value to them, 0 otherwise)	0.933 (0.25)
OTHUSEYN (1 if respondent rated the use of the Catawba River by their friends and family as an important reason why the management plan would be of value to them, 0 otherwise)	0.602 (0.49)
EXISTYN (1 if respondent rated the knowledge that water quality in the basin was being protected regardless of their use of it as an important reason why the management pan would be of value to them, 0 otherwise)	0.774 (0.42)
LIKELYYN (1 if respondent thought the management plan was somewhat or very likely to succeed, 0 otherwise)	0.778 (0.42)
ITEMYN (1 if respondent owns at least one item used for outdoor water-based recreation, 0 otherwise)	0.709 (0.45)
ENVORG (1 if respondent belonged to an environmental or conservation organization, 0 otherwise)	0.123 (0.33)
QUALWORS (1 if respondent thought water quality in their area has gotten worse over the last 5 years, 0 otherwise)	0.495 (0.50)
TAPGOOD (1 if respondent thought their tap water was above average or excellent quality, 0 otherwise)	0.467 (0.50)
AGE (age of respondent)	49.894 (14.73)
DATELAG (number of days between when the information booklet was mailed to respondent and the interview was conducted)	25.034 (20.89)
NEWAREA (1 if respondent had lived in the basin 5 years of less, 0 otherwise)	0.120 (0.33)
UNIVYN (1 if respondent somewhat or completely trusted universities, 0 otherwise)	0.707 (0.46)
EDUYN (1 if respondent had completed some college or higher, 0 otherwise)	0.615 (0.49)
SEX (1 if respondent was male, 0 otherwise)	0.544 (0.50)
INCOME (household income of respondent)	55929.01 (39333.57)

Table 2
Binomial Logit Estimation Results (NO=1)
Catawba data

VARIABLE	PARAMETER ESTIMATES	MARGINAL EFFECTS
INTERCEPT	1.060 (1.77)[a]	0.203 (1.75)
WTPAMT	1.105 (10.54)	0.211 (10.92)
TAXYN	0.626 (3.08)	0.112 (3.34)
WPCONTRO	0.141 (.73)	0.027 (.74)
USEYN	0.178 (.85)	0.034 (.85)
DRQUALYN	-0.494 (1.49)	-0.104 (1.38)
OTHUSEYN	-0.660 (3.12)	-0.130 (3.06)
EXISTYN	-0.834 (3.98)	-0.175 (3.74)
LIKELYYN	-1.080 (5.40)	-0.231 (5.06)
ITEMYN	-0.271 (1.37)	-0.053 (1.33)
ENVORG	-0.951 (2.96)	-0.149 (3.77)
QUALWORS	-0.240 (1.36)	-0.046 (1.36)
TAPGOOD	0.379 (2.15)	0.073 (2.14)
AGE	0.003 (.44)	0.001 (.44)
DATELAG	-0.002 (.47)	-0.000 (.47)
NEWAREA	-0.550 (1.90)	-0.094 (2.15)
UNIVYN	-0.665 (3.59)	-0.135 (3.42)
EDUYN	-0.436 (2.34)	-0.085 (2.30)
SEX	-0.167 (.92)	-0.032 (.91)
INCOME	-0.008 (3.08)	-0.002 (3.09)

[a]Numbers in parentheses are absolute values of t-ratios.

Table 3
LCMNL Estimation Results
Catawba data

VARIABLE	"NO" EQUATION	"YES1" EQUATION
INTERCEPT	1.549 (1.39)[a]	-3.337 (1.25)
WTPAMT	1.736 (7.04)	1.252 (-3.43)
TAXYN	0.433 (1.09)	-0.462 (0.72)
WPCONTRO	0.425 (1.21)	0.656 (-1.00)
USEYN	-1.032 (2.31)	-2.799 (4.19)
DRQUALYN	-0.143 (0.31)	1.478 (-0.97)
OTHUSEYN	0.271 (0.68)	2.073 (3.10)
EXISTYN	0.120 (0.31)	2.572 (2.44)
LIKELYYN	-1.519 (3.98)	-0.726 (1.17)
ITEMYN	0.431 (1.07)	1.688 (2.58)
ENVORG	-2.183 (4.50)	-3.534 (2.63)
QUALWORS	-0.986 (2.85)	-1.579 (2.78)
TAPGOOD	-0.295 (0.87)	-1.473 (2.50)
AGE	0.008 (0.74)	0.011 (0.60)
DATELAG	-0.006 (1.01)	-0.012 (0.88)
NEWAREA	-0.787 (1.60)	-0.362 (0.37)
UNIVYN	-0.534 (1.69)	0.332 (0.56)
EDUYN	-0.786 (-2.22)	-0.715 (1.15)
SEX	-0.871 (2.42)	-1.428 (2.44)
INCOME	-0.009 (2.04)	0.000 (0.02)

[a] Numbers in parentheses are absolute values of t-ratios.

Table 4
Marginal Effects from the LCMNL Model
Catawba Data

VARIABLE	"NO"	"YES1"	"YES2"
INTERCEPT	0.612	-0.761	0.148
	(3.23)	(1.94)	(0.39)
WTPAMT	0.276	0.095	-0.371
	(6.13)	(1.78)	(4.71)
TAXYN	0.133	-0.123	-0.009
	(2.60)	(1.20)	(0.08)
WPCONTRO	0.039	0.090	-0.129
	(0.79)	(0.85)	(1.14)
USEYN	0.004	-0.447	0.442
	(0.05)	(2.97)	(3.55)
DRQUALYN	-0.153	0.292	-0.140
	(1.52)	(1.10)	(0.67)
OTHUSEYN	-0.111	0.371	-0.260
	(1.66)	(2.57)	(2.09)
EXISTYN	-0.237	0.498	-0.261
	(3.02)	(2.51)	(1.75)
LIKELYYN	-0.272	-0.013	0.285
	(4.88)	(0.13)	(2.40)
ITEMYN	-0.044	0.285	-0.241
	(0.72)	(2.29)	(2.00)
ENVORG	-0.187	-0.492	0.678
	(1.58)	(2.18)	(3.73)
QUALWORS	-0.086	-0.219	0.304
	(1.51)	(2.28)	(3.05)
TAPGOOD	0.057	-0.255	0.199
	(1.02)	(2.32)	(1.89)
AGE	0.001	0.001	-0.002
	(0.53)	(0.51)	(0.70)
DATELAG	-0.000	-0.002	0.002
	(0.47)	(0.76)	(1.00)
NEWAREA	-0.142	-0.004	0.146
	(2.05)	(0.03)	(0.90)
UNIVYN	-0.144	0.107	0.037
	(3.22)	(1.06)	(0.35)
EDUYN	-0.113	-0.071	0.184
	(2.25)	(-0.75)	(1.65)
SEX	-0.073	-0.199	0.273
	(1.27)	(1.97)	(2.60)
INCOME	-0.002	0.001	0.001
	(3.27)	(0.53)	(0.79)
$3P_i/N$	0.322	0.255	0.423

[a] Numbers in parentheses are absolute values of t-ratios.

Table 5
Means of Independent Variables for Ten Highest Predicted Probabilities
Catawba Data

VARIABLE	MEAN for YES1	MEAN for YES2
WTPAMT (dollar amount of management plan, from $5 to $250)	82.000	23.000
TAXYN (1 if respondent rated reducing state and federal taxes important to them, 0 otherwise	0.400	0.400
WPCONTRO (1 if respondent had previously heard of efforts to control water pollution, 0 otherwise)	0.900	0.900
USEYN (1 if respondent rated their own use of the Catawba River as an important reason why the management plan would be of value to them, 0 otherwise)	0	0.900
DRQUALYN (1 if respondent rated the quality of the drinking water in their area as an important reason why the management plan would be of value to them, 0 otherwise)	1.000	0.900
OTHUSEYN (1 if respondent rated the use of the Catawba River by their friends and family as an important reason why the management plan would be of value to them, 0 otherwise)	1.000	0.700
EXISTYN (1 if respondent rated the knowledge that water quality in the basin was being protected regardless of their use of it as an important reason why the management pan would be of value to them, 0 otherwise)	1.000	1.000
LIKELYYN (1 if respondent thought the management plan was somewhat or very likely to succeed, 0 otherwise)	0.900	1.000
ITEMYN (1 if respondent owns at least one item used for outdoor water-based recreation, 0 otherwise)	1.000	0.800
ENVORG (1 if respondent belonged to an environmental or conservation organization, 0 otherwise)	0	1.000
QUALWORS (1 if respondent thought water quality in their area has gotten worse over the last 5 years, 0 otherwise)	0.400	0.900
TAPGOOD (1 if respondent thought their tap water was above average or excellent quality, 0 otherwise)	0	0.500
AGE (age of respondent)	52.300	44.200
DATELAG (number of days between when the information booklet was mailed to respondent and the interview was conducted)	16.600	33.100
NEWAREA (1 if respondent had lived in the basin 5 years of less, 0 otherwise)	0.100	0.200
UNIVYN (1 if respondent somewhat or completely trusted universities, 0 otherwise)	0.900	1.000
EDUYN (1 if respondent had completed some college or higher, 0 otherwise)	0.500	1.000
SEX (1 if respondent was male, 0 otherwise)	0.100	0.800
INCOME (household income of respondent)	63578.20	132619.40

Table 6
Variable Definitions and Summary Statistics
Saginaw Data

VARIABLE NAMES AND DEFINITIONS	MEAN (STANDARD DEVIATION)
LNBID (natural log of bid amount, from $5 to $150)	4.372 (0.69)
ACRES/1000 (number of acres purchased for preservation)	2.546 (1.42)
TRAVCOST/10 (travel cost to the Saginaw bay area)	5.031 (2.93)
SUBCOST/100 (travel cost to a substitute recreational site)	1.467 (0.47)
INCOME3/10 (household income of respondent)	5.252 (2.84)
MEMBER (1 if respondent is a member of an environmental or conservation organization, 0 otherwise)	0.406 (0.49)
LIKELY2 (1 if respondent thought the preservation plan was likely to succeed, 0 otherwise	0.477 (0.50)
Number of observations	281

Table 7
Binomial Logit Estimation Results (NO=1)
Saginaw Data

VARIABLE	PARAMETER ESTIMATES	MARGINAL EFFECTS
INTERCEPT	-2.292 (2.18)	-.565 (2.17)
LNBID	1.058 (4.94)	0.261 (4.93)
ACRES/1000	0.040 (0.41)	0.001 (0.41)
TRAVCOST/10	0.085 (1.18)	0.021 (1.18)
SUBCOST/100	-0.717 (1.35)	-0.177 (1.35)
INCOME3/10	-0.138 (2.18)	-0.034 (2.18)
MEMBER	-0.670 (2.33)	-0.165 (2.36)
LIKELY2	-1.217 (4.32)	-0.292 (4.57)

[a]Numbers in parentheses are absolute values of t-ratios.

Table 8
LCMNL Estimation Results
Saginaw Data

VARIABLE	"NO" EQUATION	"YES1" EQUATION
INTERCEPT	-1.490 (0.20)	-0.332 (0.04)
LNBID	1.631 (1.81)	0.659 (0.64)
ACRES/1000	1.931 (1.24)	2.102 (1.41)
TRAVCOST/10	0.656 (0.82)	0.592 (0.74)
SUBCOST/100	-8.608 (0.96)	-8.451 (0.95)
INCOME3/10	0.604 (0.95)	0.828 (1.30)
MEMBER	0.772 (0.55)	1.712 (1.13)
LIKELY2	-0.872 (0.87)	0.400 (0.34)

[a] Numbers in parentheses are absolute values of t-ratios.

Table 9
Marginal Effects from the LCMNL Model
Saginaw Data

VARIABLE	**"NO"**	**"YES1"**	**"YES2"**
INTERCEPT	-0.307	0.205	0.102
	(0.67)	(0.32)	(0.12)
LNBID	0.280	-0.158	-0.123
	(2.71)	(0.73)	(0.42)
ACRES/1000	0.084	0.111	-0.195
	(0.35)	0.75)	(0.50)
TRAVCOST/10	0.051	0.010	-0.062
	(0.70)	(0.18)	(0.50)
SUBCOST/100	-0.547	-0.288	0.835
	(0.54)	(0.43)	(0.50)
INCOME3/10	-0.006	0.073	-0.067
	(0.06)	(1.89)	(0.52)
MEMBER	-0.130	0.241	-0.111
	(0.63)	(2.19)	(0.48)
LIKELY2	-0.291	0.254	0.037
	(3.83)	(1.35)	(0.22)
$3P_i/N$	0.548	0.342	0.110

[a] Numbers in parentheses are absolute values of t-ratios.

Table 10
Means of Independent Variables for Ten Highest Predicted Probabilities
Saginaw Data

VARIABLE	MEAN for NO	MEAN for YES1	MEAN for YES2
LNBID	5.085	3.606	3.496
ACRES/1000	2.813	3.375	1.125
TRAVCOST/10	3.658	4.110	7.612
SUBCOST/100	0.892	1.611	1.837
INCOME3/10	2.46	7.740	5.350
MEMBER	0.000	1.000	0.100
LIKELY2	0.000	1.000	0.900

Table 11a
Means Test between LCMNL Probabilities and Follow-up Probabilities

LCMNL Probabilities

Follow-up Probabilities	Yes1	Yes2	No
Yes1	.409		
Not Yes1	.256		
Yes2		.122	
Not Yes2		.075	
No			.635
Not No			.362
t-statistic	6.53*	3.27*	10.00*

*significant at the α=.01 level.

Table 11b
Simple Logit Models Predicting the Follow-up categories

Variable	Yes1	Yes2	No
Constant	-1.99	-2.29	-2.42
LCMNL probability	3.47	2.40	4.85
LLR	37.29*	8.42*	80.81*

*significant at the α=.01 level.

References

Aadland, D., and A. J. Caplan, "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias," *American Journal of Agricultural Economics*, 85, 2, 2003, pp. 492-502.

Aitken, M., and D. B. Rubin, "Estimation and Hypothesis Testing in Finite Mixture Models," *Journal of the Royal Statistical Society, B* 47, 1985, pp. 67-75.

Berndt, E. R., B. H. Hall, R. E. Hall, and J. A. Hausman, "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3, 1974, pp. 653-65.

Blumenschein, K.,M. Johannesson, G. C. Blomquist, B. Liljas and R. M. O'Conor, "Hypothetical Versus Real Payments in Vickrey Auctions," *Economic Letters*, 56, 1997, pp. 177-180.

Blumenschein, K, M. Johannesson, G. C. Blomquist, B. Liljas and R. M. O'Conor, "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation," *Southern Economic Journal*, 65, 1998, pp.169-177.

Blumenschein, K., M. Johannesson, K. K. Yokoyama and P. R. Freeman "Hypothetical versus Real Willingness to Pay in the Health Care Sector: Results from a Field Experiment," *Journal of Health Economics*, 20, 2002. pp. 441-457.

Blumenschein, K., G.C. Blomquist, M. Johannesson, N. Horn, and P. Freeman, "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment," unpublished manuscript, University of Kentucky, June 22, 2004.

Brown, T. C., I.Ajzen, and D. Hrubes, "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation," *Journal of Environmental Economics and Management*, 46, 2003, pp. 353-361.

Caudill, S. B. "Searching for Hidden Unemployment in Transition Economies: An Approach Based on a Logit Model with Missing Information," Unpublished Working Paper, 2003.

Caudill, S. B. and P. A. Groothuis, "Modeling Hidden Alternatives in random Utility Models: An Application to "Don't Know" Responses in Contingent Valuation," *Land Economics* 81, 2005, pp. 445-454.

Caudill, S. B., M. Ayuso, and M. Guillen, "Fraud Detection Using A Multinominal Logit Model with Missing Information," forthcoming in the *Journal of Risk and Insurance*.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum, "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods" *Journal of Environmental Economics and Management* 33, 1997, pp.151-162.

Champ, P. A., and R.C. Bishop, "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias, *Environmental and Resource Economics,* 19, 2001, pp. 383-402.

Cramer, J. S. and G. Ridder, "Pooling States in the Multinomial Logit Model," *Journal of Econometrics* 47, 1991, pp. 267-72.

Cummings, R G., G. W. Harrison, and E. E. Rutström, "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review*, 85, 1995, pp. 260-266.

Cummings, R. G., S. Elliot, G. W. Harrison, and J. Murphy, "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, 105, 1997, pp. 609-621.

Cummings, R. G., and L. O. Taylor, "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method," *American Economic Review*, 89, 1999, pp. 649-665.

Dempster, A. P., N. M. Laird, and D. B. Rubin, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1977, pp. 1-38.

Eisen-Hecht, J. I., and R. A. Kramer, "A Cost-Benefit Analysis of Water Quality Protection in the Catawba Basin," *Journal of the American Water Resources Association*, April 2002: pp. 453-465.

Greene, W.H. and D.A. Hensher, "A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit," Working Paper ITS-WP-02-08, 2002.

Harrison, G. W., "Experimental Evidence on Alternative Environmental Valuation Methods," forthcoming in *Environmental and Resource Economics.*

Hausman, J. A., Abrevaya, J., and F. M. Scott-Morton, "Misclassification of a Dependent Variable in a Discrete-response Setting, *Journal of Econometrics*, 1998, 87: pp. 239-69.

Johannesson, M., B. Liljas, and P. Johansson, "An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions," *Applied Economics*, 30, 1998, pp. 643-647.

Johannesson, M., G. C. Blomquist, K. Blumenschein, P. Johansson, B. Liljas, and R. M. O'Conor, "Calibrating Hypothetical Willingness to Pay Responses," *Journal of Risk and Uncertainty*, 8, 1999, pp. 21-32.

Kramer, R. A. and J. I. Eisen-Hecht, "Estimating the Economic Value of Water Quality in the Catawba River Basin of North and South Carolina," *Water Resources Research*, September 2002: 21 1-10.

List, J. A., and C. A. Gallet, "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values: Evidence from a Meta-Analysis?" *Environmental and Resource Economics,* 20, 2001, pp. 241-254.

Loomis, J. B., A. Gonzalez-Caban, and R. Gregory, "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics,* 70, 4, 1994, pp. 499-506.

Loomis, J. B., T. Brown, B. Lucero, and G. Peterson, "Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual Willingness to Pay," *Land Economics,* 72, 1996, pp. 450-461.

Lusk, J. L., "Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice," *American Journal of Agricultural Economics*, 85, 4, 2003, pp. 840-856.

Magder, L. S., and J. P. Hughes, "Logistic Regression When the Outcome is Measured with Uncertainty," *American Journal of Epidemiology*," 146: 1997, pp.1 95-203.

Poe, G. L., J. E. Clark, D. Rondeau, and W.D. Schulze, "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation," *Environmental and Resource Economics*, 23, 2002, pp.105-131.

Rao, C.R., "Estimation and Tests of Significance in Factor Analysis," *Psychometrica* 20, 1955, pp. 93-111.

Vossler, C. A., R. G. Ethier, G. L. Poe, and M. P. Welsh, "Payment Certainty in Discrete Choice Contingent Valuation Responses: Results from a Field Validity Test," *Southern Economic Journal*, 69, 4, 2003, pp. 886-902.

Whitehead, J. C., P. A. Groothuis, R. Southwick and P. Foster-Turley, "The Economic Values of Saginaw Bay Coastal Marshes" Report 2005.