Predictive Validity in Valuation Assessment For the Provision of Farmland Amenities: A Disaggregated Approach

Julie Insignares Santos[1]

[1]PhD Candidate, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Email: jinsignares@my.uri.edu

Predictive Validity in Value Assessment For the Provision of Farmland Amenities: A Disaggregated Approach

March 28, 2014

1 Introduction

Research in environmental and natural resource valuation rely heavily on hypothetical survey data to estimate values for public goods [4, 36, 26, 29, 31]. The advantage of the stated preference approach lies in the ability to construct hypothetical scenarios in which the researcher has the ability to define the attributes of the scenarios and their levels [37]. In contrast, revealed preference experiments confine the researcher to the realm of realized behavior.

Often, however, results from hypothetical surveys have not been satisfactorily reflective of observed behavior. Data from stated preference experiments tend to overestimate actual demand, particularly in the case of public goods valuation[6, 22, 25, 23, 24]. This phenomenon is generally termed "hypothetical bias" and has been linked to: subject pool variety, differences in information provided across experiments, social norms, and whether willingness-to-pay vs. willingness-to-accept is being measured[14]. However, in light of the fact that market-based instruments for valuing ecosystem services are on the rise [12, 8, 20, 21, 10], demand for the kind of information that stated preference methods alone can offer is steadily rising. Inferring values from stated preference surveys that suffer from hypothetical bias may induce policy-makers to set policy objectives at levels that will result in inefficient outcomes.

Several studies have found significant differences between stated and revealed values for goods and services that are derived from the environment. Aadland and Caplan [1] compared stated and revealed preferences for curbside recycling programs. Brooks and Lusk [3] compared survey responses to scanner data on sales of organic and rBST-free milk. Champ and Bishop [7] utilized certainty scales to identify hypothetical bias in response to questions regarding the voluntary purchase of wind-generated energy for a period of one year. Murphy and colleagues [25] evaluated hypothetical payments and binding offers to contribute voluntarily to the Massachusetts Chapter of The Nature Conservancy using a cheap-talk script to mitigate hypothetical bias.

Concurrently, there is mounting evidence of the importance of accommodating attribute processing heterogeneity such as ignoring or 'non-attendance' to one or more features of the good and lexicographic preferences in stated and revealed preference data analysis[5, 30, 15, 16, 19, 31]. The recent literature recognizes that individuals responding to surveys often employ simplifying strategies that are not consistent with conventional random utility maximization. I posit that attribute processing rules may explain some of the observed hypothetical bias in data from surveys concerning values for public goods. The objective of this study is to identify differences between stated and revealed payments by drawing upon the latest research on attribute processing rules (APRs) in order to examine the extent to which APRs can be useful for identifying sources of hypothetical bias.

From a disaggregated perspective, I identify response strategies and test their impacts on estimated values for agricultural ecosystem services. Using a latent-class model that incorporates APRs, I am able to identify strategies in responses such as 'yea-saying' [14], attribute non-attendance (ANA) [5], and lexicography on particular attributes [19] in order to make inferences about how these dynamics affect measurements at the aggregate level. These issues are particularly relevant to valuation of ecosystem services for several reasons. First, because of the non-rival nature of public goods and services, as long as enough support is generated, non-payers cannot be precluded from consumption of the good. Thus, in a hypothetical situation, it is ad-

vantageous to send a positive signal. In addition, because such goods are not traded in conventional markets, assigning economic values for them can prove cognitively arduous. Thus, employing simplifying heuristics when making choices might be more likely than for private goods for which monetary valuation comes more naturally. I contribute to the growing literature on the importance of accommodating different attribute processing rules by incorporating process heterogeneity in the analysis of a choice experiment regarding farmland ecosystem services and testing the validity of the results against revealed market demand. To my knowledge, this is the first work to examine attribute processing strategies in hypothetical markets for public goods and their implications for realized behavior in the field.

In contrast to studies that utilize survey measures to identify and calibrate hypothetical responses (i.e. cheap-talk scenarios and certainty scales), the approach used in this analysis can be used in the absence of such measures. I utilize stated preference choice experiment (CE) and revealed preference market experiment (RP) data from a project involving the design and implementation of a local market-like process for ecosystem services in the community of Jamestown, RI. This particular experiment was uniquely conducive to analysis at the level of the individual for several reasons. Both CE and market experiments were administered to the same sample of individuals in a small community in rural Rhode Island. Thus, differences in subject pool (i.e. university students vs. grocery store shoppers) can be ruled out. In addition, a substantial amount of demographic and attitudinal information was collected regarding the respondents. This information is used to make assessments of feasible processing strategies. In contrast to most studies using stated and revealed preference data, this project conducted a choice experiment before the market experiment. Finally, care was taken to engineer the market good and the mechanism of exchange to be as closely consistent with the hypothetical choices as possible. Moreover, nearly identical elicitation mechanisms including the pivotal mechanism and provision point mechanisms were administered in order to reduce freeriding. This way, I am able to examine the performance of different elicitation mechanisms to address free-ridership as a separate issue.

The paper proceeds as follows. Section 2 describes the ecosystem service under analysis and provides a brief review of the CE and RP experiments and the hypotheses to be tested. Section 3 describes the method used to combine the data. Section 4 presents results and section 5 provides a summary and conclusion.

2 Constructing a local market for wildlife preservation

The empirical application involved establishing a market for wildlife protection marketable to the members of the rural communities surrounding local farmland. Every spring, hay farms on Jamestown, Rhode Island serve as nesting grounds for a species of ground nesting birds with a large migratory range and charismatic song called the Bobolink (*Dolichonyx oryzivorus*). Historically, hay fields in many U.S. states have been in decline as preference is given to other crops. In addition, crops are cut 2-3 weeks earlier than has been historically (since the 1940s and 1950s) [28]. This shift in cropping practices has led to serious mortality for Bobolink fledglings. Consequently, Bobolinks are now protected under the Migratory Bird Treaty Act and are listed as a Species of Special Concern in some states. The Bobolink experiment sought to transfer compensatory payments from community members to farmers to delay harvesting of hay crops thereby permitting Bobolink offspring to fledge and avoid devastation at the hands of the plow.

The two phases of the experiment were administered as a mail-in survey and solicitation to the inhabitants of Jamestown, RI and spanned a three-year period from 2006 to 2008. In all treatments, the target mailing was all deliverable addresses in the town. The SP survey was designed as a multi-question choice experiment (CE) experiment mailed to the residents of Jamestown from October to December of 2006. There were 5 questions comparing two potential contracts and a sixth question with one potential contract. The sixth response was not utilized in this analysis. Each contract was described by a list of attributes (see Table 1). There were six attributes described: 1. acreage under contract to delay harvest (Acres), 2. the number of acreage to restore to active farmland (Restore), 3. whether the acreage was found to have a high or low concentration of Bobolink (HighBobolink), 4. whether the

contracted acreage is viewable from the road (View), 5. whether or not a birdwalk is offered (Tour), and 6. the cost of implementing the contract. Respondents were presented with two competing contracts displayed side by side. Individuals were then asked whether they would choose contract A, contract B, both, or neither. That individuals were given a choice to choose Both contracts is a novel feature of the SP survey that permits identification of yea-sayers in the sample. A full description of the survey design and implementation can be found in Uchida et al.[35].

The SP mailing was comprised of five sections in total. In addition to the CE task outlined above, there were three additional sections that elicited opinions with respect to values of farmland amenities, rural character preservation, community attachment, and the importance of fairness in payment for services provided by farmland amenities. This last line of questioning was meant to assess the impact of different elicitation mechanisms on the decision to participate in the market. Several public goods payment mechanisms were administered in order to test in their field effectiveness at mitigating free-riding. A summary of the relevant findings of the attitudinal sections is listed in Table 2. The last section of the mailing collected demographic information from the respondent.

The second phase of the project was a revealed-choice experiment whereby acceptable contracts were drawn up between the mediators (researchers at the University of Rhode Island in association with EcoAsset Markets, Inc.) and farmers in the same community of Jamestown, RI. Community members were solicited for payment toward provision of the contracts, again via mail-in solicitation. This phase of the project, marketed as The Nature Services Exchange of Jamestown, was open to the residents of Jamestown in early 2007 and again in early 2008. The markets successfully provided five of ten potential field contracts.

The SP and RP treatments were designed to be as closely consistent as possible. However, there were some differences across surveys. A comparison of the attributes across both phases of the project is listed in Table 3. The market experiment did not offer participants a choice to restore fallow land to active cultivation. Also not

included in the market process was an invitation to an expert-led birdwalk. Plot support for Bobolink activity was represented differently across treatments. In the CE experiment, the attribute "Low/High Bobolink Concentration" was mildly correlated with plot size and, in all specifications, was found to be insignificant. The market experiment included two separate measurements of potential to support Bobolink populations. The number of Bobolink territories observed in 2006 was used in the 2007 treatment and the number of Fledglings expected to be supported by the field was used in 2008. Both CE and market treatments described contracts as having a view or no view, with the addition of a 'partial view' option in the 2007 market treatment. The field size attribute had a broader range in the stated preference survey and costs were comparable across years.

Market participation, as expected, was higher for individuals who returned the SP survey. Of the 791 respondents who returned the SP survey, 764 and 713 respectively were included in the market mailing in 2007 and 2008. The individuals who chose the "Both" option consistently in the SP experiment are singled out as "yea-sayers". There are two specific issues with this group. First, choosing the Both option for each question offers no information about one's relative preferences for the attributes of the contracts. Specifically, no trade-off between cost and attribute levels is observed. Thus, a model that assigns either arbitrarily low values for the marginal utility of income or arbitrarily high values for the marginal valuation of all attributes is likely to result. These respondents will likely pull the population estimates up. Of the original set of CE respondents, there were 109 (13.8%) individuals who chose the Both option for all choice situations.

The second issue with this subset of the SP respondents is the actual signal that is being sent by adopting a "yea-saying" strategy. Caudill and colleagues [6] present evidence that yea-sayers come in two varieties: that some of the respondents are truly more interested in and willing to pay for ecosystem preservation but that others may not ultimately be willing to pay the stated amount. This second group may simply be sending a signal that farmland amenities are important to them without expending

the mental effort to assess whether they would actually be willing to pay the stated amount. Because of the hypothetical nature of the survey, there is no consequence to this type of behavior and this type of behavior is generally termed "hypothetical bias".

Over all, individuals who returned the CE survey were more likely to return the market experiment solicitation, make an offer, and offer higher bids. The subset of yea-sayers, in fact, have even higher participation and offers (with the exception of the 2008 treatment). However, the vast majority of respondents to the CE experiment stated that they would participate in at least one of the scenarios offered them. Given this fact, the participation rate for these individuals is lower than one would expect. There were two possible reasons for lower participation rates in the 2008 treatment of the market experiment. First, the recession by then was fully in place and, second, there had been some controversy involving the land trust's efforts to purchase conservation easements on three farms in Jamestown. This may have had the effect of generally reducing confidence in projects to support farmland. A summary of market experiment participation separated by year and by major group (SP respondents, SP yea-sayers, and RP response only) is listed in Table 4. Based on this information, I test several hypotheses.

First, I exploit techniques from burgeoning research on attribute non-attendance (ANA) to test outcomes for individuals who had low sensitivity to the contract attributes that did not transfer to the market experiment. Attribute non-attendance involves ignoring one or more attributes when comparing alternatives in a choice scenario. I model attribute non-attendance to the Restore and Tour attributes and hypothesize that these estimates of stated WTP are in a sense more reliable because they assume that the decision process is more aligned with the RP scenario.

Hypothesis 1: Non-Attendance to attributes that were not included in the market experiment yields more consistent estimates of WTP.

Next, I examine the market behavior of yea-sayers in this context. Because of the unique presence of the Both option, yea-saying is easily detectable in the SP

application. Based on the discussion above, I examine the behavior of this class of individuals in the market in order to determine the extent of hypothetical bias inherent in yea-saying. I then compare the incidence of hypothetical bias in this group against the rest of the sample.

Hypothesis 2: There is both higher incidence of hypothetical bias and higher revealed payments linked to yea-saying behavior.

The overarching goal of this analysis is to explore whether accounting for attribute processing rules impacts the predictive validity of SP data. While the SP survey required substantially more cognitive effort and costs in terms of investment of time, the RP survey traded these costs for actual monetary commitments. Therefore, while both experiments involved costs, they were of different types. Indeed, Hensher and Greene [16] suggest a link between attribute processing rules and hypothetical bias, implying that failure to accommodate for APRs might significantly contribute to what has been termed hypothetical bias in the literature. Thus, I examine a model of response that accounts for preference heterogeneity alone against one that incorporates attribute processing rules. I hypothesize that models that incorporate APRs outperform models based on random utility maximization in terms of both model fit and predictive validity.

Hypothesis 3: Behavioral outputs of SP measures that account for APRs predict payments in experiments involving real payments with greater accuracy than models that account for preference heterogeneity alone.

3 Methods

The analysis proceeds in four steps. First, two competing models of SP response are estimated: a latent class model (LC) which incorporates taste heterogeneity and a latent class model that also accommodates attribute processing rules (LC-APR). Then, in order to make inferences about how certain processing strategies manifest in revealed preference experiments, the individual-specific conditional probabilities of class membership from the LC-APR model are utilized in a model of market participation for the 2007 and 2008 market experiments. Conditional on market participa-

tion, an offer equation is used to examine differences in offer amounts by respondent type, controlling for contract and demographic covariates. Specific attention is paid to classes that model non-attendance to attributes that were omitted from the RP experiment. I test whether there is evidence that non-attendance to the Restore and Tour attributes leads to more reliable estimates of market outcomes. Finally, predictive validity of the LC and LC-APR models is compared.

The LCL Model

In order to identify strategic behaviors that might violate the assumptions of neo-classical utility maximization and evaluate the predictive performance of a model that incorporates these strategic behaviors, I utilize an approach that is commonly used in the absence of direct survey queries about response rules. The Latent Class Logit (LCL) model with restrictions for APRs is particularly useful for this kind of analysis. The LCL model has been used both to explore patterns of attribute non-attendance and other violations of continuous preference ordering as well as modeling non-parametric preference heterogeneity. There is by now a substantial literature which uses the LCL model to identify attribute processing rules (APR), especially in the absence of self-reported non-attendance[5, 30, 15, 16, 19]. Most cite improvements in model fit and more realistic estimates of WTP when attribute processing strategies are incorporated in this manner. The LC model is widely exploited in marketing and transportation studies but has recently been used in cases of public goods valuation. Several studies cite the importance of accommodating APRs in choice modelling and there is growing evidence that modeling APRs improves model fit and leads to estimates of marginal WTP that are more consistent [31, 17, 15, 30]. If an attribute is ignored, then relative trade-offs that involve that attribute are not meaningful. That is, no increase/decrease in the ignored attribute compensates for a change in an attended attribute. This is particularly concerning if the attribute being ignored is the Cost attribute as WTP estimates cannot be calculated.

The first-stage model combines Train's [34] Expectation Maximization algorithm for nonparametric estimation of the random parameter latent class logit model with

Hess et al.'s [18] expansion of attribute non-attendance for heterogeneous taste variation. I model non-attendance to key CE survey attributes that were omitted in the market experiment: acres of restored farmland and invitation to a bird walk. In addition, I aim to catch and contain the yea-sayers in the sample, whose insensitivity to contract price would otherwise inflate marginal values for all other attributes. Attribute non-attendance is expected to be a significant problem with this particular type of choice task since respondents are not likely to be familiar with the ecosystem service for offer and thus may make unforeseen assessments of the true meaning of the attributes of the contract, or to decide that a particular attribute is too cryptic to assess a value for.

The Expectation-Maximization algorithm applied to latent class modelling has been utilized by Train [34] as a form of non-parametric estimation of underlying taste heterogeneity whereby a discrete distribution whose accuracy in approximating the true underlying distribution rises with the number of parameters. This is an extension of Bhat [2] where increasing the number of classes allows for better approximation of taste heterogeneity. Several authors have noted advantages of the LC model over the popular Mixed Logit Model in capturing taste heterogeneity [11, 32, 18].

The Latent Class specification proceeds as follows. Given the standard choice modeling scenario, N agents choose among J alternatives in each of T choice occasions, let y_{njt} be an indicator variable equal to 1 if agent n chooses alternative j in choice situation t. Each alternative is defined by a set of attributes with varying levels and the choice of attributes is assumed to result from standard neoclassical random utility maximization. Further assume that there are C distinct sets of taste parameters in the population, $\beta = \{\beta_1, ..., \beta_C\}$. A set of C Multinomial Logit models represent the C discrete support points for the distribution of tastes in the population. In this framework, there are two sets of unknowns: the β_cs, that is, the estimates of taste parameters for each class, and the class membership status of the agents. The number of classes, C, is chosen by the researcher based on measures of fit. If agent n belongs to class c, then the probability of observing her choices is the product of logit formulas

over all choice situations. A set of C conditional logit models weighted by class share are repeatedly estimated. The weights are constructed via fractional multinomial logit. One of the advantages of the LC model is the ability to model heterogeneity in preferences via the class membership model. By including respondent demographics one can model heterogeneity in preferences without the need to interact demographics with the attributes of the choice situation. The class membership model is estimated simultaneously via fractional multinomial logit.

Model fit is generally assessed based on minimizing an information criterion such as AIC, BIC, or CAIC [5, 31, 27]. If the information criteria do not agree on which model is preferred, the researcher must choose based on examination of standard errors and feasibility of parameter signs. Because the EM algorithm does not involve maximizing the likelihood function, special attention must be paid to assessing local vs. global maximum attainment. This is achieved by testing several starting points to ensure that a global maximum has been obtained. Each candidate model was estimated from fifty random starting points. When the number of classes is relatively small, variation in BIC was relatively low from one estimation to the next. The variance in BIC rose with the number of classes. From previous analysis, it is quite likely that preferences for the contract attributes are highly correlated. Hess et al. [18] point out that latent class models incorporate this correlation inherently through class membership probabilities.

The LC-APR model used in this analysis is fundamentally different from conventional LC models in which classes are not representative of specific behaviors[5]. For this reason, Hensher et al. [15] refer to the model as a "probabilistic decision process model" whereby the class membership probabilities represent the probability of a typical respondent exhibiting the behavior modeled in a class. Therefore, the model specification search is undertaken in a different manner. The assumptions about processing strategy are outlined first. That is, the model structure is defined. Then, the appropriate restrictions are imposed on each class defining the response strategy and then the model is estimated. In this case, the primary response behaviors of interest

are non-attendance to the attributes that were left out of the market experiment and yea-saying.

For guidance on choosing likely APRs, I summarize key findings from the attitudinal section of the SP survey. Four of the five sections of the stated preference survey consisted of several questions that asked participants to rate statements about their opinions regarding their community, farms and wildlife, and farmland amenities. Most of the questions required respondents to rate the statements on a five point Likert scale from Strongly Agree to Strongly Disagree. The statements pertained to the attributes of the contracts that the participants would subsequently be comparing and were worded as follows: "Open space in agricultural use is important to me as part of Jamestown" and, "It is important to me that I can view birds and other wildlife when I walk near farms". Nearly 97% of respondents indicated that they agreed or strongly agreed that open space in agricultural use was an important feature of their community. Undeveloped woodland was also found to be important to most of the respondents. However, the responses were mixed with regard to whether maintaining remaining agricultural landscapes was more or less important than maintaining undeveloped woodland. Participants were also asked whether they would join an expert-led bird walk if invited. There was a mix of responses to this question. A summary of these findings can be found in Table 2. Overall, the attitudinal findings imply that non-attendance to the Restore and Tour attributes is a distinct possibility in the data.

The convention with regard to using latent class models with APR restrictions is to first identify candidate APRs so that they can be tested for inclusion. In many cases, the parameters are constrained to be equal across classes and attribute non-attendance is specified by restricting the parameter to equal zero in a particular class. The rationale for constraining parameters across classes is to focus on attribute non-attendance without concern for preference and scale heterogeneity. I am interested in accommodating taste heterogeneity as well as process heterogeneity. However, doing so complicates the analysis quite a bit as the combinations of potential behaviors

rises exponentially. Thus, to simplify the analysis, I focus on non-attendance to attributes that were included in the SP experiment but left out of the RP treatments. While doing so facilitates tests of Hypothesis 1 above, it is also somewhat reinforced by the attitudinal findings listed above. That is, there is some qualitative evidence that individuals would not participate in a guided birdwalk if offered, and that some respondents might be ambivalent with regard to restoring fallow land to active cultivation. In addition, I let the taste heterogeneity be guided by evidence from the full-attendance classes. That is, I first tested one class for each type of APR. A layer of taste heterogeneity was added by increasing the number of full attendance classes until the lowest information criteria were obtained. I then added a layer of taste heterogeneity on to the ANA classes by assuming two of each ANA class. An outline of this process can be found in Table 5. The Cost ANA and All attribute ANA were confined to one class each for all specifications. The reason for not testing higher dimensions on these APRs was that constructing estimates of WTP is inconvenient for the All ANA class and practically impossible for the Cost ANA class.

For comparison against a model without APRs, I perform the classical LC model specification search by testing up to thirteen unrestricted classes. The unrestricted model is used as a baseline for comparison of performance against a model that incorporates APRs.

Constructing Individual-level WTP estimates

In order to assess the predictive validity of the LC-APR model, it was necessary to construct individual estimates of WTP for the contract that was presented to the individuals for each year of the RP market experiment. Individual-specific conditional probabilities of class membership conditional were used to weight the within-class parameters which were then applied to the standard formula for calculation of WTP based on Hannemann's formula [13]. The procedure is outlined below.

Given the K x C class parameter estimates, I estimated the conditional probabilities of each individual belonging to each of the classes. I use these conditional probabilities as weights on the class parameters to construct the marginal utilities for

each individual[11]. I then combined the estimated parameters with the RP contract attributes. I constructed choice-instance expectations of WTP by combining the individual constructed marginal utilities of acres and view, subtracting the status quo estimate, and dividing by the marginal utility of income.

I utilize within and across class variance-covariance to construct individual-specific estimates of WTP as follows:

$$WTP_{RP,year} = \frac{1}{R}\sum_{r=1}^{R}\sum_{c=1}^{C}\eta_{cn}^{r}(\beta^{r},\theta^{r}) \times \frac{\beta_{A,c}^{r}Acres_{RP,year} + \beta_{V,c}^{r}View_{RP,year} - \beta_{No,c}^{r}}{-\beta_{cost,c}}$$
$$= \frac{1}{R}\sum_{r=1}^{R}\sum_{c=1}^{C}\eta_{cn}^{r}(\beta^{r},\theta^{r}) \times \widehat{SPWTP}_{RP,year,c} \quad (1)$$

I generate 1,000 random draws of a multivariate normal distribution utilizing the parameters and variance-covariance matrix from the LC-APR model. For each draw, r, the individual-specific conditional probability of class membership are calculated given the parameter values, $\eta_{cn}^{r}(\beta^{r},\theta^{r})$, which are dependent upon the class membership parameters θ as well as the K x C β parameter estimates. The class membership probabilities weight the expected WTP for the contract administered to the individual in a given year. The mean and 95% confidence intervals for the 1,000 draws of the individual-level values are used to analyze consistency in WTP across SP/RP treatments.

Analyzing Market Responses

I examine market participation by APR via Random Effects Probit model of the decision to return the market mailing. Building upon previous analysis [33], an individual is considered to participate in the market if she returned the market mailing, even if her offer amount was zero. I test for differences in participation rates among SP vs. RP respondents in general, and class membership probability specifically, controlling for elicitation features and other demographic characteristics. The class membership probabilities are constructed from the conditional probabilities of class membership for individuals who returned the CE survey and from unconditional probabilities of class membership based on the results of the class membership model

described above and given the particular respondent's demographic characteristics.

Contingent upon market participation, I quantify differences in actual offers based on class membership, contract characteristics of the market mailing, and demographic characteristics. A Selection-Adjusted Interval Regression econometric specification is used to account for the mixture of discrete choice and payment card data. The selection adjustment is achieved via calculation of Inverse-Mills ratios (IMRs) described in Swallow et al[33] and based on Wooldridges [38] panel version of a Heckman-type selection model.

Finally, I analyze whether stated WTP are consistent with RP payments by constructing a set of rules that define "consistency" for this data. The set of criteria determine whether an individual responded consistently, or whether they under- or over-valued the market good based on the estimates of stated WTP. The criteria depend on the type of payment format, discrete choice (DC) or open-ended (OE), the respondent received.

I define consistent behavior in this context as follows. First, if the offer amount was within the 95% confidence interval of estimated willingness to pay from the choice experiment (SPWTP) for individual i, then she has made an offer that is consistent with her choices in the CE survey. There is one other case in which I consider behavior to be consistent. That is, if the lower bound of the 95% confidence interval on $\widehat{SPWTP}$ is above the highest dollar value in the range of the open-ended treatment or the point value in the discrete choice treatment, and that value is chosen as the offer amount, I consider this behavior to be consistent.

Inconsistent choices are of two types. If behavior is not found to satisfy the two previous criteria, then the SP estimates overvalue the market contract if the lower bound of the 95% CI is greater than the offer amount. I identify these individuals as exhibiting hypothetical bias. Alternately, if the upper bound of the 95% CI is below the offer amount, the respondent is considered to have under-valued the good in the SP scenarios.

4 Results

Descriptive Statistics and Class Allocation

Upon preliminary examination of the SP experiment data itself, there appeared a few clear strategies. For instance, the yea-sayers (those who answered that they would purchase both contracts for all questions) comprised a substantial share of the overall respondents (13.8%). There was also evidence that many individuals were simply choosing the lowest cost option (10.7%). These qualitative findings were used to examine the performance of the LC-APR model.

The final LC-APR model was arrived at after methodical testing. The best model in the specification search, based first on BIC and then on highest average maximum conditional class membership probability had five classes: four restricted classes and one full attendance class (Table 5). The final five-class LC-APR model included the following APRs: Restore and Tour ANA, Tour ANA, All ANA, Cost ANA, and one full attendance (unrestricted) clas.

The LC-APR model results suggest that full attendance was not a majority strategy (Table 6). Individuals are assigned to classes based on highest conditional probability of membership. There were 173 respondents (22%) for whom full attribute preservation was a best fit. All attributes in the full attendance, Restore/Tour attribute non-attendance, and Tour non-attendance classes are significant and have the expected signs. The All attribute non-attendance class and Cost ANA class parameters are insignificant. Individuals who fit these classes with high conditional probability act in a way that makes it difficult for preference trade-offs to be calculated with accuracy.

Individuals who chose both contracts for all choice experiment questions (the yea-sayers) fit the Cost ANA class with extremely high conditional probability (at least 90%). The remaining seven individuals in the Cost ANA class made only one choice that was not the Both alternative. The unfortunate drawback of this class is that constructing willingness to pay estimates involves division by the cost parameter which is restricted to zero. Therefore, I found that the estimates of WTP for these

individuals, using the method described in Equation 1, produced wildly high negative or positive values because the cost parameter at the individual level was exceeding close to and equally likely to fall on either side of zero. In addition, confidence intervals around these estimates ranged in the thousands.

Following some authors, I included a class that represented full attribute Non-Attendance (All ANA) to capture idiosyncratic behavior. There were 82 respondents who fit this classification best. A portion of these respondents were found to have extreme reactions to a change from the referendum vote to one of the elicitation mechanisms.[1] Fourteen of the respondents (17%) were "protest votes", that is, they chose the No Buy option for all choice occasions. The remainder of the All ANA class exhibited seemingly random response behavior. The parameter estimate for the Cost parameter for this class is positive and very close to zero, leading the average WTP to be negative and drastically large.

There were 420 individuals who fit the Restore/Tour ANA and Tour ANA classes best. Recall that the primary goal of modeling a Retore/Tour ANA class is to test whether individuals who were relegated to this class made choices that were more consistent in the market experiment, which did not include these contract features. This class had a high positive value on the View attribute relative to the other classes, perhaps implying lexicography on this contract characteristic and indicated that a segment of the population is highly concerned with preserving the aesthetics of unharvested farmland. An alternative interpretation is that, if a parcel can be viewed from the road, then ensuring that farmers uphold the contract is possible. Monitoring compliance on a parcel that cannot be seen from a road would be difficult. In the membership model of the LC-APR, membership to this class was comprised of individuals who had high values of the Equality variable (Table 8). This variable was constructed from responses to the analysis of the survey responses regarding preferences for fair and equal payments for ecosystem services.

[1]The SP experiment sample was split into two main groups. Group 1 was issued CE treatments in which the respondents made choices under two separate scenarios: a referendum vote and the prescribed elicitation mechanism. Group 2 was not administered the referendum and thus made all five choices given a single elicitation mechanism

The second class (Tour ANA only) had low but positive values for the attended attributes. Participants who consistently chose the cheaper alternative fit this class best. Ideally, the LC-APR model would have relegated lexicographic preferences on cost to an All ANA class. This fact lends support to a further refinement in which an indicator variable for the lower cost alternative and lexicography on this variable might be included as a class. The refinement was not implemented here because it would lead to a rise in the number of individuals for which estimates of WTP would not be possible. In essence, it is likely not accurate to assume that individuals who fit this class with high probability were in fact ignoring the Tour attribute but that instead these individuals have overall low values for the farm-wildlife contract features. A summary of the classes with average WTP for an average sized parcel with a view is listed in Table 7.

For comparison, an unrestricted Latent class model was estimated (Table 9). The six class unrestricted model succeeded in relegating the yea-sayers to a single class (Class 3) as well. This feature highlights the advantage of the LC model as a non-parametric representation of taste heterogeneity. If conventional mixed logit estimation was performed, then the commonly exercised assumption of normally distributed coefficients means that the population-level estimates of the parameters would be pulled up by the influence of these responses.

Market Experiment Outcomes

Participation

The Random Effects Probit model of market participation (Table 5) measured the effects of demographics, contract characteristics, and first-stage class membership on the decision to return the market experiment survey. An individual is considered to have participated in the market experiment if she returned the payment card, even if the amount offered was zero. Consistent with previous analysis, significant demographic determinants of participation were the respondent's age, donation history, and a history of mail-order from children's catalogues. Age and a history of donation both increase the likelihood of returning the market solicitation. Mail-ordering from

children's catalogues decreases the likelihood of participation, presumably because household resources of time and money are more limited for this group.

Of the contract attributes found to influence participation, only the log of the minimum amount of the solicitation was found to be significant. As in Swallow et al[33], I acknowledge that this implies that at least some of the individuals did open the solicitation and look at it before deciding not to return the mailing. None of the effects-coded variables for mechanism treatment were found to contribute to the decision to participate in the market experiment.

In order to investigate participation given the first stage SP results, I included the conditional probability of class membership for each of the respondents estimated from the LC-APR model and the unconditional probabilities of class membership for the market experiment respondents who did not return a CE survey. These measures capture the probability of membership into an APR class, conditional on CE choices if available and unconditional probabilities for respondents who returned the market experiment only. Again, the unconditional probabilities are influenced by the LC class membership model which takes account of demographic attributes (Table 8). Consistent with the qualitative findings (Table 5), having returned the choice experiment survey significantly increased the likelihood of returning the market mailing. The full attendance class was omitted for identification purposes. Negative and significant parameter estimates on the class membership variables implies that individuals who were likely to attend to all attributes in the SP survey had the highest likelihood of participation in the market experiment relative to the other classes. Interestingly, high probability of membership in the third class (All ANA) yields the lowest participation rate. This class included all individuals who chose the No Buy option for all questions and so this finding is fairly consistent with expectations given the LC results. Individuals who belong with high probability to the Cost ANA class were not found to have a participation level that was significantly different that the full attendance class. As these individuals mostly chose the highest level of contribution in the SP survey, this result is not surprising.

Offers

Contingent upon participation in the market, the selection-adjusted interval regression of offer amount results indicate that class membership and contract payment format had important impacts on the amount of money offered in payment toward the farm-wildlife contracts (Table 5). Individuals who received a discrete choice (DC) solicitation on average made much higher offers than those who were issued low range open-ended (OE) payment cards ($10 - $80). Those who were issued the higher range payment cards ($35 - $120) contributed at a level intermediate to the low OE and DC treatments. This finding, robust to alternate specifications, is consistent with findings from previous studies regarding differences between payment card and discrete choice elicitation mechanisms. Other aspects of the contracts themselves were not found to significantly impact offer amounts.

Most of the class membership regression coefficients have significant explanatory power . The parameter estimates are to be interpreted as differences from the set of individuals who fit the full attendance class with high likelihood. Relative to full attendance, all LC-APR classes are found to offer lower amounts in the market experiment. According to these results, the order of offers from highest to lowest was

$$FullAttend > CostANA > TourANA > AllANA > Restore/TourANA \qquad (2)$$

compared to what the LC-APR predicts

$$CostANA > FullAttend > TourANA > Restore/TourANA > AllANA \qquad (3)$$

Interestingly, the respondents who fit the Cost ANA class best (yea-sayers) were found to make lower offers than individuals who were likely to attend to all attributes. Because the cost range was nearly identical across CE and market experiment treatments, the level of hypothetical bias in this group is evident. That is, if yea-sayers were to behave in a manner consistent with their responses in the hypothetical survey, one would expect these individuals to choose the highest possible offer amounts in

revealed preference situations. In fact, the results show that, while the average offer is high for this group, it is not the highest. Thus, if full attendance implies that these individuals invested more in terms of cognitive effort, this investment manifests as higher WTP in the market.

The Restore and Tour ANA individuals revealed lowest payments in the market experiment but also lowest positive values in the choice experiment. Therefore, this result is not particularly surprising. Similarly, members of the Tour ANA class had low stated values for all attended attributes, but, were not found to consistently choose the No Buy option. Thus, it follows that these individuals also exhibit low revealed values.

Measuring Hypothetical Bias and SP undervaluation

Based on the rules of consistency outlined above, I compare the LC-APR model against the LC model with preference heterogeneity alone (Tables 12 and 13). The average value of the hypothetical bias for the LC attribute processing model was \$37.69 compared to \$52.79 for the unrestricted model. This points to clear evidence that the LC APR model performs better at mitigating the effects of hypothetical bias (Hypothesis 3). Both models produced similar results with regard to undervaluing the market good. Of the respondents who were found to undervalue the market good in the SP treatment, 3 chose the lowest value on the payment card and about half were issued a discrete choice solicitation. This may exemplify the advantage that the DC format maintains over the open-ended format: that some individuals may feel pressured to participate and will accept a discrete choice payment out of this compulsion.

Focusing on the results of the attribute processing model gives a sense of possible sources of hypothetical bias. The highest proportion of consistent values can be found for the class of individuals who fit the Restore and Tour ANA class best. This supports the first hypothesis of this research in finding that individuals who are found to be rather insensitive to the attributes that do not transfer to the market can be expected to produce more consistent or reliable estimates of willingness to pay. In

addition, the Cost ANA class contained the highest proportion of respondents who exhibited hypothetical bias. Because calculating a realistic estimate of WTP for this class was impossible, consistency for the Cost non-attendance class was defined as choosing the highest level on the open-ended treatment, accepting the discrete choice price, or writing in an amount above the range in both cases. Under this definition of consistency, 67% of yea-sayers exhibited hypothetical bias. Of the consistent responses for this group, each individual chose the highest value presented them. This is 1.5 times the amount of hypothetical bias exhibited by the full attendance class which had the next highest proportion. Therefore, I find some support for Hypothesis 2, that yea-sayers reveal a high level of hypothetical bias but also exhibit high valuation otherwise.

5 Summary and Conclusions

Stated choice experiments are good tools for assessing the nature of potential demand for a new product. The strength of the discrete choice experiment lies in the ability of the researcher to generate data that contains sufficient variation so as to best infer the nature of the trade-offs among attributes that potential market participants will make. However, unlike new product development in the realm of tradable goods, a significant challenge to the success of eliciting preferences for nonmarket goods is inducing respondents to reveal their true values for a nonexclusive public good.

Behavioral economists and choice modelling researchers have found mounting evidence that the neoclassical model of utility maximization does not always succeed in characterizing economic behavior. Attribute processing strategies have an effect on valuation in CE surveys and that these processing strategies have implications for market participation and contribution in this unique dataset. It has been suggested that accounting for such strategies may help alleviate some of the well-documented differences in measured stated vs. revealed values, especially for public goods.

The study outlined in this paper offered a unique means by which to examine differences between revealed and stated preferences for two reasons. First, the CE survey included an option to purchase both contracts, thereby releasing the respondent from

the need to make any trade-offs. Individuals who were inclined to choose this option were found to be of two types: respondents with high values for the contracts and their attributes, and respondents who exhibited hypothetical bias. By comparing the behavior of these respondents in the subsequent market experiment, I was able to clearly observe the level of HB resultant from this strategy of hypothetical response. My findings indicate that yea-sayers may be expected to exhibit high contribution levels, but overall, were not found to participate to a larger extent than other respondents. This implies that some means of partitioning this group into individuals whose actual value for the good is high and those who are engaging in yea-saying would assist in determining how much weight to assign these responses when making inferences about public goods values.

In addition, the CE survey involved the presentation to respondents of a 'good' that was not previously available to the community. Rather than testing a new dimension of an already existing consumption good (i.e. an added label to an existing carton of milk or a new transportation option), the survey involved assigning value for a new product with possibly unfamiliar characteristics. Therefore, applying simplifying rules to make the choice easier was a distinct possibility. Indeed, I find that a model that incorporates APRs succeeds in mitigating the upward bias of SP responses.

The method employed in this research can be useful for many applications. In general, future research may utilize these findings to test for shortcomings of the survey design or in identifying and calibrating for specific processing heuristics so that a more accurate assessment of values can be derived from hypothetical survey data. With the growing emphasis on market-based solutions for ecosystem services provisions, it is increasingly important to accurately measure values and identify challenges at the market design stage. This research provides some preliminary evidence for what works with regard to getting the prices right.

Table 1: Summary of Choice Experiment contract attributes

Attribute	Description	Levels
Acres	Number of acres to be placed under contract upon which farmer will delay mowing and harvesting	10, 25, 40, 55
Restore	Number of acres to be restored to active hay fields, not restricted to delayed mowing/harvesting	0, 10, 20, 30
High Bobolink	Level of expected fledglings saved (correlated with acreage)	Low, High
View	Whether the parcel will be viewable from the road	View, No View
Tour	Whether individuals paying into the contract are invited to an expert-led birdwalk	Tour, No tour

Note: There were five defining attributes of each of the CE contracts. The attributes and their levels are listed above. Source: [35].

Table 2: Summary of Attitudinal Findings from Supplementary CE Questions

Question	Summary of Response
Open space in agricultural use is important to me as part of Jamestown	97% Agree or Strongly Agree
Undeveloped woodland is important to me as part of Jamestown	Over 90% Agree or Strongly Agree
It is important for me to maintain remaining agricultural landscapes than to maintain undeveloped woodlands in Jamestown	Results indicate that there is some heterogeneity with respect to this issue
It is important to me that I can view birds and other wildlife when I walk near farms	Highly skewed toward Agree or Strongly Agree
Improving farm management to provide Bobolink habitat	Majority response was "somewhat valuable" with significant heterogeneity
Improving public access and educational opportunities such as walking tours to visit farms and watch wildlife	Over 79% of responses in somwhat to extremely valuable, with the rest in the somewhat to not at all

Note: The results from a selection of the attitudinal section of the Hypothetical Choice experiment are listed above. Source: [9]

Table 3: Comparison of Attributes and Levels across SP and RP treatments

Attribute	Choice Experiment	Market 2007	Market 2008
Field size to delay harvesting	10 – 55	10 – 18	10
Acres of farmland restored to cultivation	0 – 30		
Bobolink Concentration	Low, High		
# of 2006 territories		1 – 4	
# of Fledglings			6—10, 10—14, 14—18
View from road	View, No view	View, Partial, No view	View, No view
Mechanism	SP, VCM:PPMBG, PR, PM, UPA	PM, PR, UPA	PM, PR, UPC
Cost	\$10 — \$105	\$10 — \$120	\$10 — \$120

Note: The mechanisms were as follows: SP: referendum for tax increase, VCM:PPMBG: provision point mechanism with money-back guarantee, PR: provision point with money-guarantee and proportional rebate of excess funds, PM: pivotal mechanism, UPA: uniform price auction, UPC: uniform price cap.

Table 4: Summary of Market Response

	Total		SP respondents		SP Yeah-sayers		RP only	
	2007	2008	2007	2008	2007	2008	2007	2008
N	2791	2680	764	713	105	101	2027	1967
Returned Market mailing	13%	8%	30%	16%	34%	14%	7%	5%
Made Offer	7%	5%	19%	11%	27%	11%	3%	3%
Average Value of Offer	$47.94	$46.49	$50.30	$46.98	$60.54	$41.82	$42.11	$45.83

Note: The table is divided into four groups: all individuals who returned the market mailing, those who returned both the market and CE surveys, the subset of CE respondents who chose the "Both" option for all questions, and all other individuals who were mailed a market experiment survey.

Table 5: Specification Search

	Number of Classes								
Non-Attendance	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Restore/Tour	1	2	1	1	1	2	1	1	2
Tour	1	1	2	1	1	1	2	1	2
All	1	1	1	2	1	1	1	2	2
Cost	1	1	1	1	1	1	1	1	1
None	1	1	1	1	2	2	2	2	2
BIC	**7137**	7149	7149	7151	7142	7172	7176	7166	7249
AIC	6904	6863	6859	6875	6847	6826	6825	6829	**6805**

The specification search is outlined above. Each column represents the number of classes specified in each model to have the parameter restrictions that imply the non-attendance patterns listed in the first column. Boldfaced values represent the best performing information criterion.

Table 6: Latent Class APR Model Results: Five Classes

	Class 1 Restore/Tour	Class 2 Tour	Class 3 All	Class 4 Cost	Class 5 Full Attend
Cost	-0.074***	-0.045***	0.001	0	-0.017***
No ASC	-0.487***	-1.582***	-1.168	-8.718	-0.871***
Acres	0.011**	0.028***	0	0.106	0.02***
Restore	0	0.026***	0	0.13	0.055***
View	0.305***	0.124**	0	8.95	0.211***
Tour	0	0	0	0.095	0.3***
Share	0.159	0.386	0.085	0.145	0.225
BIC			7137		
AIC			6902		
LL			-3401		

Notes:*** p<0.01, ** p<0.05, * p<0.1

Table 7: Summary of Classes

Class	Definition	Unique Features	Avg. WTP for 10 acres with view	N
1	Restore & Tour ANA	Possible Lexicography on View	$12.19	120
2	Tour ANA	Includes individuals who chose cheaper contract consistently	$44.13	300
3	All ANA	Protest Votes, Extreme reactions to payment mechanism, quixotic behaviors	$-1168.00	82
4	Cost ANA	Includes all yea-sayers	NA	116
5	Full Attendance	High Values	$75.41	173

Note: N is the number of individuals for whom the maximum conditional probability places them in the class

Table 8: Latent Class APR Membership Model

Variable	Class 1 Restore/Tour	Class 2 Tour	Class 3 All	Class 4 Cost
Age	0.01	0.01	0.018	0.004
Income	-0.004	-0.001	-0.000	-0.002
Equality	0.929***	0.186	-0.130	-0.290
Mail Order Kids	-0.256	0.05	-2.42	-0.170
Education	-0.549***	-0.311***	-0.934	-0.112
Gender	-0.397	-0.05	-1.07	-0.353
Constant	2.45**	1.79***	3.50	0.251

Notes:*** p<0.01, ** p<0.05, * p<0.1
Class 5 is omitted from estimation for purposes of identification.

Table 9: Latent Class, Unrestricted Model

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Cost	-0.049***	-0.016***	-0.126***	-0.094***	-0.015***	-0.007
No ASC	-1.537***	0.107	-7.783	-0.651***	-1.057***	-2.175***
Acres	0.033***	-0.032***	1.869***	0.01***	0.037***	0.008
Restore	0.024***	0.024*	-0.187***	0.034***	0.075***	0.03***
View	0.1*	0.96***	-0.609	0.297***	0.221**	0.124
Tour	0.094*	0.072	7.531	-0.224**	0.241**	0.279***
Share	0.36*	0.038***	0.131	0.015	0.156	0.166
BIC	7096					
AIC	6904					
LL	-3411					

Notes:*** p<0.01, ** p<0.05, * p<0.1

Table 10: Participation Equation

	Coeff	z
Constant	-2.881***	-5.380
Contract Attributes		
Acres	0.00423	0.261
View	0.111	1.086
# Bobolink Territories	0.0137	0.279
# Fledglings	0.003	0.120
lnMinAmt	-0.0982*	-1.849
Discrete Choice	0.0874	1.133
Mechanisms		
Uniform Price Auction	0.0639	0.697
Pivotal Mechanism	0.0342	0.530
Uniform Price Cap	-0.128	-1.123
Class Membership		
Returned SP	1.212***	11.54
$P(RestoreTourANA \mid CE)$	-0.454*	-1.864
$P(TourANA \mid CE)$	-0.369*	-1.700
$P(AllANA \mid CE)$	-0.987**	-2.442
$P(CostANA \mid CE)$	-0.152	-0.619
$P(FullAttend \mid CE)$	–	–
Other Demographics		
Log Purchasing Power	-0.0982*	-1.849
Age	0.0167***	4.184
Donation History	0.260***	2.585
Environmental Donations	0.135	0.976
Mail Order Kids Mags	-0.304***	-2.782
Year 2008	-0.459	-1.639
$\ln\sigma_u^2$	0.366**	2.191
LL	-1579	
N	5415	
χ^2	180.38	
p	0.000	

*** p<0.01, ** p<0.05, * p<0.1

Note: The class membership variables represent the conditional probability of class membership given the individual's responses to the choice experiment survey.

Table 11: Payment Equation

	Coeff	z
Constant	21.99	0.269
Contract Attributes		
Discrete Choice	39.62***	5.197
OE High	18.37**	2.471
Acres	0.0351	0.0327
View	8.249	1.212
# Bobolink Territories	-3.511	-1.089
# Fledglings	2.504	1.443
Mechanisms		
UPAe	-9.901	-1.601
PMe	4.541	0.988
UPCe	-1.399	-0.180
Class Membership		
Returned SP	22.58***	3.773
$P(RestoreTourANA \mid CE)$	-91.21***	-5.786
$P(TourANA \mid CE)$	-45.64***	-3.363
$P(AllANA \mid CE)$	-60.67**	-2.369
$P(CostANA \mid CE)$	-26.97**	-1.991
$P(FullAttend \mid CE)$	–	–
Other Demographics		
LogPurchasing Power	4.18	0.775
Age	0.0225	0.0603
Donation History	-1.605	-0.209
Environmental Donations	7.825	0.833
Missing Donation Information	22.47	1.154
Mail Order Kids Mags	-31.29***	-3.849
Inverse Mills and 2008 Indicator		
2008	-36.58	-0.775
IMR	4.518	0.129
IMR 2008	3.699	0.126
σ_u	46.36***	12.10
σ_e	23.60***	5.937
LL	-740.63	
BIC	1652.88	
N	576	
χ^2	94.11	
p	0.000	

*** p<0.01, ** p<0.05, * p<0.1

Note: The class membership variables represent the conditional probability of class membership given the individual's responses to the choice experiment survey.

Table 12: RP Offer Evaluation - LC APR Model

	Consistent Values	Hypothetical Bias	Under Valued	Total
Restore/Tour ANA	31	3	12	46
Tour ANA	52	49	28	129
All ANA	13	4	1	18
Cost ANA	20	35	0	55
Full Attendance	31	36	17	84
Total	147	124	58	332

Note: Individuals are relegated to a class based on the highest conditional probability of class membership. Consistent responses indicate that the individual's offer lies within the confidence interval of their estimated WTP from the choice experiment. Hypothetical bias exists when the lower bound of an individual's estimated WTP lies above the market offer amount. Similarly, under valuing occurs when individuals state a lower WTP in the choice experiment than offered in the market experiment.

Table 13: RP Offer Evaluation - Unrestricted Model

	Consistent Values	Hypothetical Bias	Under Valued	Total
Class 1	50	48	31	129
Class 2	2	5	2	9
Class 3	20	31	0	51
Class 4	32	3	11	46
Class 5	27	23	4	54
Class 6	13	21	9	43
Total	144	131	57	332

5 References

[1] David Aadland and Arthur J Caplan. Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. *American Journal of Agricultural Economics*, 85(2):492–502, 2003.

[2] Chandra R Bhat. An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1):34–48, 1997.

[3] Kathleen Brooks and Jayson L Lusk. Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *American Journal of Agricultural Economics*, 92(4):1229–1241, 2010.

[4] Trudy Ann Cameron and Jeffrey Englin. Welfare effects of changes in environmental quality under individual uncertainty about use. *Rand Journal of Economics*, pages S45–S70, 1997.

[5] Danny Campbell, David A Hensher, and Riccardo Scarpa. Non-attendance to attributes in environmental choice analysis: a latent class specification. *Journal of environmental planning and management*, 54(8):1061–1076, 2011.

[6] Steven B Caudill, Peter A Groothuis, and John C Whitehead. The development and estimation of a latent choice multinomial logit model with application to contingent valuation. *American Journal of Agricultural Economics*, 93(4):983–992, 2011.

[7] Patricia A Champ and Richard C Bishop. Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. *Environmental and Resource Economics*, 19(4):383–402, 2001.

[8] Roger Claassen, Andrea Cattaneo, and Robert Johansson. Cost-effective design of agri-environmental payment programs: Us experience in theory and practice. *Ecological Economics*, 65(4):737–752, 2008.

[9] Christopher Anderson Crystal Fry, Emi Uchida and Stephen Swallow. Summary report: Working farms and wildlife in jamestown, ri: 2006 survey of jamestown residents. Technical report, University of Rhode Island, Kingston, RI, 2007.

[10] Erik Gómez-Baggethun, Rudolf De Groot, Pedro L Lomas, and Carlos Montes. The history of ecosystem services in economic theory and practice: from early notions to markets and payment schemes. *Ecological Economics*, 69(6):1209–1218, 2010.

[11] William H Greene and David A Hensher. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698, 2003.

[12] Maryanne Grieg-Gran, Ina Porras, and Sven Wunder. How can market mechanisms for forest environmental services help the poor? preliminary lessons from latin america. *World development*, 33(9):1511–1527, 2005.

[13] W Michael Hanemann. Welfare evaluations in contingent valuation experiments with discrete responses. *American journal of agricultural economics*, 66(3):332–341, 1984.

[14] David A Hensher. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological*, 44(6):735–752, 2010.

[15] David A Hensher, Andrew T Collins, and William H Greene. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. *Transportation*, pages 1–18, 2012.

[16] David A Hensher and William H Greene. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical economics*, 39(2):413–426, 2010.

[17] David A Hensher, John M Rose, and William H Greene. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates

and a warning for stated choice experiment design. *Transportation*, 39(2):235–245, 2012.

[18] Stephane Hess, Moshe Ben-Akiva, Dinesh Gopinath, and Joan Walker. Advantages of latent class over continuous mixture of logit models. Technical report, working paper, 2011. URL http://www. stephanehess. me. uk/papers/Hess_Ben-Akiva_Gopinath_Walker_May_2011. pdf, 2011.

[19] Stephane Hess, Amanda Stathopoulos, and Andrew Daly. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3):565–591, 2012.

[20] Nicolas Kosoy, Miguel Martinez-Tuna, Roldan Muradian, and Joan Martinez-Alier. Payments for environmental services in watersheds: insights from a comparative study of three cases in central america. *Ecological Economics*, 61(2):446–455, 2007.

[21] Timm Kroeger and Frank Casey. An assessment of market-based approaches to providing ecosystem services on agricultural lands. *Ecological Economics*, 64(2):321–332, 2007.

[22] John A List and Craig A Gallet. What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, 20(3):241–254, 2001.

[23] Jayson L Lusk and F Bailey Norwood. Bridging the gap between laboratory experiments and naturally occurring markets: An inferred valuation method. *Journal of Environmental Economics and Management*, 58(2):236–250, 2009.

[24] Jayson L Lusk and F Bailey Norwood. An inferred valuation method. *Land Economics*, 85(3):500–514, 2009.

[25] James J Murphy, Thomas Stevens, and Darryl Weatherhead. Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? *Environmental and Resource Economics*, 30(3):327–343, 2005.

[26] Laura Nahuelhual, Maria L Loureiro, and John Loomis. Using random parameters to account for heterogeneous preferences in contingent valuation of public open space. *Journal of Agricultural and Resource Economics*, pages 537–552, 2004.

[27] Daniele Pacifico et al. lclogit: A stata module for estimating latent class conditional logit models via the expectation-maximization algorithm. *The Stata Journal, Forthcoming*, 2012.

[28] A. Poole. The birds of north america online, August 2005.

[29] Bill Provencher, David J Lewis, and Kathryn Anderson. Disentangling preferences and expectations in stated preference analysis with respaondent uncertainty: The case of invasive species prevention. *Journal of Environmental Economics and Management*, 2012.

[30] Riccardo Scarpa, Timothy J Gilbride, Danny Campbell, and David A Hensher. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European review of agricultural economics*, 36(2):151–174, 2009.

[31] Riccardo Scarpa, Mara Thiene, and David A Hensher. Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter? *Land economics*, 86(4):817–839, 2010.

[32] Junyi Shen. Latent class model or mixed logit model? a comparison by transport mode choice data. *Applied Economics*, 41(22):2915–2924, 2009.

[33] Stephen K Swallow, Christopher M Anderson, and Emi Uchida. The bobolink project: Selling public goods from ecosystem services using provision point mechanisms. Technical Report 16, Zwick Center For Food and Resource Policy, 2012.

[34] Kenneth E Train. Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1):40–69, 2008.

[35] Emi Uchida, Christopher M. Anderson, and Stephen K. Swallow. Marketing ecosystem services from agricultural land: Stated preferences over payment mechanisms and actual sales of farm-wildlife contracts. 2007 Annual Meeting, July 29-August 1, 2007, Portland, Oregon TN 9955, American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association), 2007.

[36] Christian A Vossler, Joe Kerkvliet, Stephen Polasky, and Olesya Gainutdinova. Externally validating contingent valuation: an open-space survey and referendum in corvallis, oregon. *Journal of Economic Behavior & Organization*, 51(2):261–277, 2003.

[37] John C Whitehead, Subhrendu K Pattanayak, George L Van Houtven, and Brett R Gelso. Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science. *Journal of Economic Surveys*, 22(5):872–908, 2008.

[38] Jeffrey M Wooldridge. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132, 1995.