Preference Elicitation Methods and Valuation Implications: Experimental Evidence from Valuation and Purchase of Water Quality Credits

Pengfei Liu*

pengfei.liu@uconn.edu

Department of Agricultural and Resource Economics,

University of Connecticut, 1376 Storrs Road, Unit 4021, Storrs, CT, 06269

*Corresponding Author

Stephen K. Swallow

stephen.swallow@uconn.edu

Department of Agricultural and Resource Economics and

Center for Environmental Sciences and Engineering

University of Connecticut, 1376 Storrs Road, Unit 4021, Storrs, CT, 06269

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Preliminary Draft.

Do Not Cite without Permission.

Preference Elicitation Methods and Valuation Implications: Experimental Evidence from Valuation and Purchase of Water Quality Credits*

Pengfei Liu† Stephen Swallow‡

May 26, 2016

Abstract

This paper compares different preference elicitation methods used in choice experiments. We implemented four different methods to elicit individuals' preference for a non-market good. Our four treatments include (1) a hypothetical referendum, (2) a real referendum lacking incentive compatibility, (3) a real choice with incentive compatibility and (4) a hybrid approach that combines (2) and (3). We develop a method to estimate the percentage of strategic choices in each treatment. We find that in the hypothetical referendum, about 75% to 92% individuals truthfully reveal their preference and choose the option that gives the highest utility in a choice question. Adding policy consequentiality (e.g., the real referendum) and payment consequentiality (e.g., the hybrid approach) could increase the percentage of individuals truthfully reveal their preference.

Keywords: Stated Preference, Policy Consequences, Incentive Compatibility, Welfare, Water Quality Trading

JEL: Q56, Q57, C72

*Corresponding Author, Pengfei Liu, pengfei.liu@uconn.edu.

†Department of Agricultural and Resource Economics, University of Connecticut, 1376 Storrs Road, Unit 4021, Storrs, CT, 06269

‡Department of Agricultural and Resource Economics and Center for Environmental Sciences and Engineering, University of Connecticut, 1376 Storrs Road, Unit 4021, Storrs, CT, 06269.

1 Introduction

This paper is motivated by the gap between conducting stated preference analyses and applying stated preference results for economic (especially environmental) policies involve the valuation of non-market goods. We focus on the criticisms toward the stated preference analysis for its lack of incentive compatibility. In this paper, we study the incentives for individuals to make choices in choice experiment under the utility maximization framework. We find that the stated preference method is comparable to a complicated voting system. Democratic societies have widely adopted various versions of a public voting system for presidential elections. The stated preference, or the contingent valuation method, is similar a voting system and should be recognized as an important input to public policies. Therefore, a careful economic analyses of different methods used in eliciting public preference using contingent valuations is critical for its generalizability.

Early stated preference studies focus on estimating the willingness to pay to a non-market good based on respondents' choices, such as the hunters' values for hunting permits (Bishop and Heberlein, 1979). Environmental economists started to pay attentions to the potential hypothetical bias arise from self-reported survey responses at the very begining. Different elicitation methods, including the comparisons with revealed preference data through the travel cost method are investigated in Bishop and Heberlein (1979). Hanemann (1984) provides a rigorous theoretical framework to calculate the welfare measures implied by the fitted model from discrete choice data. Opaluch et al. (1993) apply a contingent choice survey to the noxious facilities siting problem based on a constructed utility index for each options. The policy makers then could rely on the "scoring mechanism" that aggregates public reference to evaluate and select noxious facility site, possibly in combination with other technical or political consideration (Swallow et al., 1994). Stated preference method has been applied to address important environmental issues of national interest such as the *Exxon Valdez* oil spill in 1989. However, the estimated damage can range widely from \$3.8 million to \$4.9 billion. Thus, even though stated preference method can be potential power tool to assist better policy making, the use of the stated preference has been controversial (Diamond and Hausman, 1994; Hausman, 2012).

One of the major challenges is to understand the incentives issues in the stated preference studies. Simply using the self-reported survey responses as the preference elicitation method can lead to imprecise or biased coefficients estimates and potentially undermine the robustness of affected policy. For example, respondents simply do not have the incentive to state their most preferred options (the option gives them the highest utility) when choices are purely hypothetical, or respondents will not be influenced by their own choices thus the survey is inconsequential (Carson and Groves, 2007). Even when respondents have the incentive to choose the best option among a set of alternatives, and their choices are consequential in the sense that their utility could be impacted by a new policy based on their choices, ambiguous or uncertainty about how their choices could influence the policy may enable respondents to strategically choose the second best or other options. One example is the choice of second best candidates in the majority-voting rule when voters have to decide among three candidates. The consequences of such strategic incentives are overlooked in the stated preference literature. In this paper, we incorporate the possibility of strategic responses in the choice experiment and quantitatively assess the percentage of the strategic behaviors (or misrepresentatin of true preference due to insufficient incentive) using different preference elicitation methods.

Stated preference methods, such as the choice experiments, use public surveys to present realistic scenarios (which could be implemented by a future policy) and require a choice among two or more multi-attribute alternatives (Louviere et al., 2000; Adamowicz et al., 1998). Recent literature (Carson and Groves, 2007; Carson et al., 2014; Vossler and Evans, 2009; Vossler et al., 2012) focuses the attentions on whether respondents have the (economic) incentives to answer choice questions and whether the choices could produce valid measures of willingness to pay for non-market goods. Results show, in part, that respondents may depart from stating their true preferences over alternatives in a single choice scenario if (i) surveys involve a series of questions or (ii) surveys (otherwise) give respondents an impression that their answers may affect more than one decision by, say, a public agency. Economists have long understood that surveys bearing at least a minimal degree of consequentiality is a necessary condition to obtain valid preference information, with these recent contributions evaluating the role of consequentiality relative to response on each choice question (Kafle et al., 2015; Boxall et al., 2009; Day et al., 2012; DeShazo and Fermo, 2002; Holmes and Boyle, 2005; Meyerhoff et al., 2014; Zhang and Adamowicz, 2011). These progresses mitigate concerns from an otherwise purely hypothetical survey instrument. However, the connection between individual choices and policy maker's choice is not clearly spelled until very recently. Vossler et al. (2012) provide a theoretical foundation and establish conditions where individuals choices can be considered incentive compatible. However, even though individuals are motivated to answer the survey truthfully, their responses may not be very helpful to inform a policy decision since the existing incentive compatible preference elicitation methods often rely on implementing the choice from one individual.

While the literature rightly points out the difficulty of producing a survey in which choices are incentive-compatible, which would align response incentives with a respondent's true preferences. Applied valuation studies, however, frequently strive to support a variety of agency or government decisions (Opaluch et al., 1993; Swallow et al., 1992). The recent focus on the incentives surrounding a single survey question has deflected attention from using the choices for the policy input to some extent. In reality, costs of data acquisition commonly compel researchers to present a series of choices to respondents, rather than a single choice. Moreover, applied valuation surveys typically address a complexity of attribute bundles provided by, for example, agro-ecosystem management choices, and leave respondents with incomplete information regarding the portfolio of real choices that decision-makers will face, the number of choices to be made, and the distribution of preferences in the population against which a respondent could, in theory, play a strategic game. The respondent's opportunity to gain through strategic misrepresentation may be narrowly circumscribed by numerous dimensions of uncertainty (Kawai and Watanabe, 2013; Choi et al., 2011), possibly triggering a response based on cognitive dissonance (Akerlof and Dickens 1982; cf., Akerlof 1983), which may lead to truthful statements of preference in order to resolve the cognitive challenge (or psychological discomfort) while minimizing the risk of sending ineffective or even harmful signals to decision-makers. Our paper quantitatively assess the gap (i.e., the percentage of strategic choices) when subjects have the strategically opportunities compared to a situation when subjects have incentives to report preferences truthfully.

Most of the mechanisms studied to ensure incentive compatibility are debatable (Horowitz, 2006), especially when we consider potential uncertainty and risk (Karni and Safra, 1987). Mitani and Flores (2014) find that the probability of payment has a negative effect on contributions to a public good, which implies that a downward estimation bias can still exist if only one or some respondents are randomly chosen and accountable for their choices (a comment technique to assure, in theory, incentive compatibility). Harin (2014) suggests the random-lottery incentive system used in many experiments is not truly incentive compatible under a more robust model of how utility- maximizing participants in experiments react to uncertainty and the (often low) probability of having one of their choices be pivotal to a real decision within an experiment. Harin's theoretical analysis is consistent with the Kawagoe and Mori (2001) examination of the pivotal mechanism (PM) for funding public goods. The PM is a form of Clarke (1971) tax, which is incentive-compatible, but only weakly: it is a dominant strategy for an individual to truthfully reveal his or her preference in a choice, but the strategy is only weakly dominant (i.e., telling the truth never lowers the person's benefit-level, so truthfulness is incentive compatible, but failing to tell the truth does not always reduce benefits). Kawai and Watanabe (2013) find that only a small fraction (1.4% to 4.2%) of potential voters will vote strategically (e.g., vote for an alternative other than the one they most preferred). Esponda and Vespa (2014) find the about 78% subjects are nonstrategic even after possible learning in the sequence of voting experiments. If these results could be generalized to stated preference valuation studies, incentive compatibility might be less problematic in terms of its overall influence on the ability of decision-makers supported by stated preference valuation studies supporting decisions intended to enhance social efficiency. In our paper, we find that the non-utility-maximizing choices account for 8% to 25% of the total responses in a hypothetical referendum treatment and the non-utility-maximizing choices account for less than 10% of the total responses when the treatment is both policy and payment consequential though lacking incentive compatibility. Furthermore, our model is able calibrate coefficient estimates and obtain more precise estimation of utility parameters with the addition of new choice data, even though they are not produced from an incentive compatible treatment.

The problem of only relying on incentive compatible elicitation method is that the policy maker is often unable to implement the policy since the possibility to influence multiple outcomes could distort individual incentives. Methods used to assure incentive compatibility for single questions may lead the stated preference survey to deviate from the original motivation of common applied studies: soliciting preferences to rank potential policy options of public interest. For example, one academic approach to achieve incentive compatibility might tell respondents that one survey question will be chosen for implementation according to a single respondent's choice; the respondent will be chosen at random, but, clearly, a single individual's (a selected "dictator's") preference is insufficient to reflect the public preference and thus the resulting policy option implemented may be far from the social optimum use of available resources. This paper compares the estimated utility parameters from different preference elicitation methods and we find a certain degree of consistency across different treatments. The coefficient estimates from a treatment that is treatment that is both policy and payment consequential (though lacking incentive compatible) is closest to the coefficients estimated from an incentive compatible treatment.

This paper is organized as follow. Section 2 provides an analytical framework to analyze the choice incentives, similar to Vossler et al. (2012). Section 3 introduces the experiment background and procedure. Section 4 discuses the econometric model and analyze the estimation results. Section 5 explores the valuation implications from different preference elicitation methods. Section 6 concludes.

2 Model

Our analytical framework is built on Vossler et al. (2012) and extended to include a cost sharing rule. Consider a choice experiment where an individual has to complete multiple choice questions. Individual i's choice can be denoted by a vector $\mathbf{V}_i = (V_{i1}, V_{i2}, ..., V_{iJ})$. After observing individuals' choices, the policy maker chooses a policy which involves a vector of attribute $\mathbf{A}_n$ and a cost C_n, from a total of N possible policies. The set of potential policies is denoted as $\mathbf{N} = (\mathbf{A}_n, C_n)$. We assume the cost C_n for implementing the project N is solely determined by the project attributes $\mathbf{A}_n$. Thus, we have $\mathbf{N} = (\mathbf{A}_n, C_n(\mathbf{A}_n)) = \mathbf{A}_n$.

Each participant chooses $\mathbf{V}_i$ to maximize the utility

$$U_m(N(\mathbf{V}_i, \mathbf{V}_{-i}; G), T(\mathbf{V}_i, \mathbf{V}_{-i}; G); \mathbf{X}_m),$$

where $N(\cdot)$ is the policy implementation rule and $T(\cdot)$ is the cost sharing rule (such as the "payment vehicles"), both of which are functions of the individual decisions. The vector $\mathbf{V}_{-i}$ is the choices made by other participants except i, $\mathbf{X_i}$ represents individual i's individual characteristics, such as an income variable y_i; G represents the factors that influence the policy function others than individuals choices, such as policymaker's preference, cost sharing rule among individuals and realistic constraints. We use c_i to denote the cost incurred to individual i based on the cost sharing rule $T(\mathbf{V}_i, \mathbf{V}_{-i}; G)$ and have $c_i = T(\mathbf{V}_i, \mathbf{V}_{-i}; G)$. When the policy N is implemented, all the relevant individuals are affected by the same policy N and potentially benefit from the outcome determined by the project attributes $\mathbf{A}$. In this way, individual i's utility is $U_i(\mathbf{A}, c_i; \mathbf{X}_i)$ similarly for individual i' with a utility $U_{i'}(\mathbf{A}, c_{i'}; \mathbf{X}_{i'})$.

2.1 Participants' Choices

Individual i makes a sequence of decisions $\mathbf{V}_i = (V_{i1}, V_{i2}, ..., V_{iJ})$ when presented with J choice sets. Assume that the participant i make the J choices independently and maximizes individual utility $U_i(\cdot)$, which is a function of the policy maker's choice on $N(\cdot)$ and cost sharing rule $T(\cdot)$. Thus,

$$\begin{aligned} \mathbf{V}_i^* &\in argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\ &= argmax \quad U_i(N(\mathbf{V}_i, \mathbf{V}_{-i}; G), T(\mathbf{V}_i, \mathbf{V}_{-i}; G); \mathbf{X}_i), \end{aligned} \tag{1}$$

where $\mathbf{A}$ is a vector of attributes brought by a potential policy and c_i is the cost incurred by the participant i, which is *different* from the cost of the proposed project C. For the individual choices on the choice question j,

$$
\begin{aligned}
V^*_{i,j} &\in \ argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\
&= \ argmax \quad U_i(N(V_{i,j}, \mathbf{V}_{i,-j}, \mathbf{V}_{-i}; G), T(\mathbf{V}_i, \mathbf{V}_{-i}; G); \mathbf{X}_i),
\end{aligned} \tag{2}
$$

When each individual only makes on single choice, the above equation becomes:

$$
\begin{aligned}
V^*_{i} &\in \ argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\
&= \ argmax \quad U_i(N(V_i, \mathbf{V}_{-i}; G), T(\mathbf{V}_i, \mathbf{V}_{-i}; G); \mathbf{X}_i).
\end{aligned} \tag{3}
$$

When the individuals' multiple choices are independent and only one choice can influence the policy outcome (such as through a random lottery rule), the above equation becomes,

$$
\begin{aligned}
V^*_{i,k} &\in \ argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\
&= \ argmax \quad U_i(N(V_{i,k}, \mathbf{V}_{-i}; G), T(V_{i,k}, \mathbf{V}_{-i}; G); \mathbf{X}_i).
\end{aligned} \tag{4}
$$

When the policy implemented only depends on the choice from one individual (i.e., the dictator), the above equation becomes,

$$
\begin{aligned}
V^*_{i,k} &\in \ argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\
&= \ argmax \quad U_i(N(V_{i,k}; G), T(\mathbf{V}_i; G); \mathbf{X}_i).
\end{aligned} \tag{5}
$$

2.2 The Policymaker's Choice

Define the policy function $\mathcal{F} : \mathcal{V} \times \mathcal{G} \to \mathcal{N}$, where $\mathcal{V}$ is the observed/revealed choice space, $\mathcal{G}$ is the policymaker's preference space and $\mathcal{N}$ is the potential *implementable* policy space. Thus, the "optimal" policy $N^* = \mathbf{A}^*, N^* \in \mathcal{N}$ is determined by the observed choice $\mathbf{V}^* = (\mathbf{V}^*_1, \mathbf{V}^*_2, ..., \mathbf{V}^*_M)$ and the realization of policymaker's preference G.

2.3 Incentive Compatibility

In our context, incentive compatibility means that for any given choice set, participant i will always choose the option that yields that highest utility from the set of alternatives, given the policy function $\mathcal{F}$. Specifically, for the choice set k, where the available choices are $\mathbf{d}_k = (1, 2, 3, ..., D)$, incentive compatibility requires that participant i's optimal choice $V^*_{i,k}$ for question k also maximizes the utility among options available for question k. Thus,

$$
V^*_{i,k} \ \in \ argmax \quad U_i(\mathbf{a}(V_{i,k}), c_i(V_{i,k}); \mathbf{X_i}), \forall k \tag{6}
$$

where $\mathbf{V^*_i} = \{V^*_{i,1}, V^*_{i,2}, ..., V^*_{i,D}\}$. The $\mathbf{a}$ and c_m are the policy/project attributes and private cost faced by individual i in question k. Note that the policy attributes $\mathbf{a}$ in this case is only influenced by individual i's choice in question k, and will not be influenced by individual i's choices $\mathbf{V}_{i,-k}$ in other questions, other individuals' choices $\mathbf{V}_{-i}$ and the policy makers' preference G.

Compare equation (2):

$$\begin{aligned} V_{i,k}^* &\in argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\ &= argmax \quad U_i(N(V_{i,k}, \mathbf{V}_{i,-k}, \mathbf{V}_{-i}; G), T(\mathbf{V}_i, \mathbf{V}_{-i}; G); \mathbf{X}_i), \end{aligned}$$

with equation (6) on individual i's choice on question k:

$$V_{i,k}^* \in argmax \quad U_i(\mathbf{a}(V_{i,k}), c_i(V_{i,k}); \mathbf{X_i}), \forall k$$

The above two equations show potential inconsistence between the incentive compatibility and individual utility maximization objectives. When $K = 1$ and each participant is presented one choice experiment. The utility maximization behavior leads to

$$\begin{aligned} V_i^* &\in argmax \quad U_i(\mathbf{A}, c_i; \mathbf{X}_i) \\ &= argmax \quad U_i(N(V_i, \mathbf{V}_{-i}; G), T(V_i, \mathbf{V}_{-i}; G); \mathbf{X}_i). \end{aligned}$$

The incentive compatibility constraint becomes:

$$V_i^* \in argmax \quad U_i(\mathbf{a}(V_i), c_i(V_i); \mathbf{X_i}),$$

Definition 1 *A mechanism is incentive compatible if*

$$U_i(N(V_i, \mathbf{V}_{-i}; G), T(V_i, \mathbf{V}_{-i}; G); \mathbf{X}_i) \geq U_i(N(V_i', \mathbf{V}_{-i}; G), T(V_i', \mathbf{V}_{-i}; G); \mathbf{X}_i)$$

where

$$V_i \in argmax\, U_i(\mathbf{a}(V_i), c_i(V_i); \mathbf{X_i}).$$

For example, when subjects m face three options, i.e., $V_i = [A, B, N]$, where A and B are two alternative policies while N stands for the status quo. If the option A gives the highest utility among the three options, then an incentive compatible mechanism implies that,

$$U_i(N(A, \mathbf{V}_{-i}; G), T(A, \mathbf{V}_{-i}; G); \mathbf{X}_i) \geq U_i(N(B, \mathbf{V}_{-i}; G), T(B, \mathbf{V}_{-i}; G); \mathbf{X}_i)$$

and

$$U_i(N(A, \mathbf{V}_{-i}; G), T(A, \mathbf{V}_{-i}; G); \mathbf{X}_i) \geq U_i(N(N, \mathbf{V}_{-i}; G), T(N, \mathbf{V}_{-i}; G); \mathbf{X}_i).$$

The above characterizations are helpful to analyze the incentive properties of a preference elicitation method.

3 Experiment Background and Treatments

We work with the EPRI's Ohio River Basin Trading Project to implement our research design by linking tradable water quality credits with the valuation of water quality credits and co-benefits from individuals participated in our experiment. Our research focuses on enhancing knowledge of key factors affecting the use of stated preference studies for non-market valuation, in our context, the valuation toward water quality credits associated with different co-benefit profiles. The Ohio River Basin Trading Project (EPRI 2010, 2014) is an interstate trading program involves the states of Ohio, Indiana, and Kentucky. Water quality credit producers (mostly farmers) get credits awarded by reducing their loadings of total nitrogen (TN) or total phosphorus (TP) below historic practices. Besides documenting the water quality credits, the project also keeps track of many kinds of co-benefits provided by different agricultural best management practices. The co-benefits documented are verified by a third party and EPRI records ancillary benefits such as carbon sequestration, habitat enhancement, and excessive run-off, among others. The Project's online credit trading registry provides information on the implemented water quality improvement projects and associated co-benefits. The choice experiment scenarios are based on a data set composed from the online credit trading registry with the assistance of the EPRI team. Table 1 shows the definition of co-benefits and other attributes (such quantity, cost) used in the choice experiment. Figure 1 shows an example of the choice scenarios individuals face in the experiment. Liu and Swallow (2016) study the implications of incorporating co-benefits produced with water quality BMPs into the credits markets. This study explores the valuation implications in the presence of strategic behaviors or insufficient incentives in choice experiment when the preference elicitation method is considered *not* incentive compatible.

Four treatments are used in the experiment, which differ in the choice on the policy function and the payment/cost sharing rule. Treatment 1 describes hypothetical scenario where each subject has an equal probability of being paid a fixed amount independent of his or her actual choices, which we call the Hypothetical Referendum (HR). Subjects were told that their choices would be used to provide input to the U.S. Department of Agriculture and can potentially impact the future policy in the water quality credit trading market. Subjects were entered into a lottery where one subject would receive a $20 cash reward. In Treatment 1 using the Hypothetical Referendum (HR), subjects were told that:

"Your decisions in this survey will be used to provide input to the U.S. Department of Agriculture to influence future policy and decision making in Water Quality Trading markets. This could influence the future conduct of real markets like EPRI's Ohio River Water Quality Trading program."

Treatment 2 describes a policy consequential scenario where individuals' decisions would collectively influence the final policy chosen, in our context, direct our purchase decisions among various types of water quality credits. We call Treatment 2 the Real Referendum without Incentive Compatibility (RR) mechanism as subjects were told that their choice would influence real purchasing decisions based on a statistical model of how the group of participants in their treatment made choices. Each subject in Treatment 2 was also told that they would be entered in a lottery by which one subject would receive a $20 cash reward. Therefore, Treatment 2 only establishes policy consequentiality but lacks payment consequentiality. In Treatment 2 using the Real Referendum (RR), subjects were told that:

" ...We will make purchasing decisions - and spend real money to buy real water quality credits - based a statistical model of how your group made choices in this survey."

Treatment 3 specifies a consequential, incentive compatible policy function using the Independent Lottery Policy Function (Vossler et al., 2012), which we call the Real Choice with Incentive Compatibility (RC) mechanism. We implemented the random-lottery rule (Holt, 1986; Holt and Susan, 2002; Myagkov and Plott, 1997) where subjects were told that one subject will be chosen and we would implement exactly one of his or her eight decisions. For example, we will purchase the same quantity of water quality credits with an identical combination of co-benefits as implied by the chosen decision. The payment method in Treatment 3 is different from Treatments 1 and 2 in order to establish incentive compatibility. Each subject's choices have an equal probability of being chosen and the chosen subject?s payoff equals $\$20 + (\$50 - c_j)$, where the $20 is the fixed amount award used similarly as for the participation incentive in Treatments 1 and 2, c_j is the cost of purchasing the water quality credit bundle when a subject chooses option j, and $50 is the amount provided by the researchers so that a deficit for the individual is impossible. For example, if one chooses to buy a water quality bundle that cost her $30 among other alternatives, then the potential payoff from this choice is $\$20 + (\$50 - \$30) = \$40$; however, if she chose the status quo of buying nothing, the potential payoff is $\$20 + (\$50 - \$0) = \$70$. The random lottery incentive system is implemented to ensure that each choice has a positive probability of being selected as the binding outcome.[1] In Treatment 3 using the Real Choice (*RC*) with Incentive Compatibility, subjects were told that:

"...We will randomly choose one person in your group (maybe you) and execute their purchasing decision - spending real money on real water quality credits - based on a randomly drawn question from their survey."

Treatment 4 describes a hybrid approach combing the choice incentives in Treatment 2 and Treatment 3 where one's choice can be chosen randomly as a policy choice while the group's aggregated decisions will also be used for a policy choice. For example, in Treatment 4, we will implement two decisions related to a randomly chosen individual participants series of choices, one decision based on a randomly chosen question answered by that individual and a second decision based on spending an announced-and-fixed amount of money on water quality credits (with bundles of co-benefits), spending, say, $1,000 in a manner guided by a statistical model derived from the collection of choices made by members of that individual's group. We call this treatment the Real Referendum combined with Real Choice (*RR_RC*) mechanism. The *RR_RC* is a hybrid approach under which one's decision could decide individual payoff and influence purchase decision from the group decisions. Thus, in the *RR_RC* mechanism, one's choice has multiple consequences. As a result, the Treatment *RR_RC* establishes both payment consequentiality and policy consequentiality, however it is considered not incentive compatible due to one?s action could influence multiple outcomes. In Treatment 4 using the Real Referendum combined with Real Choice (*RR_RC*), subjects were told that:

"...We will randomly choose one person in your group (maybe you) and execute their purchasing decision - spending real money on real water quality credits - based on a randomly drawn question from their survey.

In addition to the decision of one individual, we will also make purchasing decisions - and spend real money to buy real water quality credits - based a statistical model of how your group made choices in this survey. "

[1] The random lottery incentive system has been widely used in the experimental economics literature and is favored due to its invariance to income effect and the ability to collect incentive compatible data within a limited budget (Wakker, 2007).

Note that Treatment 2, 3 and 4 are considered consequential since an individual's choices will influence the actual purchase decisions, while Treatment 1 is consequential only when the results obtained could influence future policy choices. In Treatment 3 and 4, the cost attribute is real and immediate under the random lottery incentive system, while in Treatments 1 and 2 the cost attribute is hypothetical in the immediate term. Only Treatment 3 is considered incentive compatible and we assume each individual chooses the option that maximizes her utility given the set of choice alternatives. Therefore, we speculate that subjects are less likely to choose the utility-maximizing option in a given set of alternatives in the hypothetical referendum in Treatment 1. Subjects tend to exhibit limited strategic choices (choose the non-utility maximizing options) in the real referendum Treatment 2, and the percentage of non-utility maximizing options could be reduced in hybrid approach in Treatment 4.

4 Empirical Modeling and Estimation Results

Based on the random utility framework (McFadden, 1973; Hanemann, 1984), individual i's utility from choosing an option j, U_{ij}, consists of an econometrically measurable, deterministic component, V_{ij}, and a random component, ϵ_{ij}, which is unobservable to econometricians and assumed to follow independently and identically distributed (i.i.d.). The measurable component V_{ij} depends on (i) the water quality credit bundle, such as quantity and the co-benefit profile X_j, (ii) the socioeconomic characteristics of the individual i, denoted by S_i and (iii) the cost c_j. Specifically, individual i's utility from choosing an option j is

$$U_{ij} = U(X_j, S_i, c_j) = V(X_j, S_i, c_j) + \epsilon_{ij}. \tag{7}$$

Given the qualitative co-benefits information (i.e., presence versus absence), we incorporate the co-benefits by interacting the co-benefits profile with the credit quantity in the utility function. This modeling strategy is more compatible with economic theory compared to treating each co-benefit as a dummy variable that additively enters the utility function. Therefore, we use a utility specification that differentiates the value of a credit by incorporating the co-benefits nonlinearly in relation to specific project characteristics. Individual i's utility function is specified as

$$V'_{ij} = Q_j^{\beta_{t,0}} c_j^{\alpha_t} exp(B_j(Z)) exp(D_i(S)), \tag{8}$$

where

$$B_j = \sum_{m=1,..,z} \beta_{t,m} Z_m \tag{9}$$

and

$$D_i = d_{ij}(\gamma_{t,0} + \sum_{m=1,..,s} \gamma_{t,m} S_m). \tag{10}$$

In the above specifications, Z_m presents the vector of co-benefit attributes and S_m represents the vector of demographic variables. For example, The vector $Z_0 = (0, 0, ..., 0)$ represents no co-benefits are associated with the water quality credit. Substitute equation (9) and (10) into equation (11) and take log transformation on both sides of the equation, we have,

$$V_{ij} = ln(V'_{ij}) = \beta_{t,0} ln(q_j) + \sum_{m=1,..,z} \beta_{t,m} Z_m + \alpha_t ln(c_j) + d_{ij}(\gamma_{t,0} + \sum_{m=1,..,s} \gamma_{t,m} S_m). \tag{11}$$

Accounting for the Scale Difference in Different Treatments We model the scale heterogeneity directly into the discrete choice model by allowing each treatment associated with a treatment-specific scale parameter σ_t (Fiebig et al., 2010; Czajkowski et al., 2014). To incorporate treatment specific scale parameter, we slightly modify the utility and get:

$$U_{ij} = U(X_j, S_i, c_j) = V(X_j, S_i, c_j)/\sigma_t + \epsilon_{ij}, \tag{12}$$

if individual i is assigned to treatment t. In the choice experiment questions, each individual considers three alternatives: options A, B involving buying a positive amount of credits or the buying nothing option N. If the individual's choice implies his or her utility is higher for alternative $j \in \{A, B, N\} \equiv J$, providing utility U_{ij} compared to all the other alternatives $U_{ik}(k \neq j, k \in J)$, then the probability that individual i chooses alternative j is calculated by

$$\begin{aligned} P_i(j) &= Pr(U_{ij} > U_{ik}, k \neq j, k \in J) \\ &= Pr(V_{ij}/\sigma_t + \epsilon_{ij} > V_{ik}/\sigma_t + \epsilon_{ik}, k \neq j, k \in J) \\ &= Pr(\epsilon_{ij} - \epsilon_{ik} > (V_{ik} - V_{ij})/\sigma_t, k \neq j, k \in J), \end{aligned} \tag{13}$$

where $Pr(\cdot)$ represents the probability operator. Based on the error structure, the probability can be simplified (McFadden, 1973) as:[2]

$$\ln P_i(j) = d_{ij} \ln \left(\frac{e^{V_{ij}/\sigma_t}}{1 + \sum_{m=\{A,B\}} e^{V_{im}/\sigma_t}} \right) + (1 - d_{ij}) \ln \left(\frac{1}{1 + \sum_{m=\{A,B\}} e^{V_{im}/\sigma_t}} \right), \quad (14)$$

where $d_{ij} = 1$ if $j \in \{A, B\}$ and $d_{ij} = 0$ if $j = N$. The treatment specific scale parameter $\sigma_t = \delta_t \sigma_1$, where σ_1 is the scale parameter for Treatment 1 and is normalized to 1 for identification purpose. Therefore, the log-likelihood function is,

$$\ln L = \sum_i \ln P_i = \sum_i \left(d_{ij} \ln \left(\frac{e^{V_{ij}/\sigma_t}}{1 + \sum_{m=\{A,B\}} e^{V_{im}/\sigma_t}} \right) + (1 - d_{ij}) \ln \left(\frac{1}{1 + \sum_{m=\{A,B\}} e^{V_{im}/\sigma_t}} \right) \right). \quad (15)$$

Let $\boldsymbol{\alpha}$=$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, $\boldsymbol{\beta}$=$(\beta_{1,0}..., \beta_{1,z}, ..., \beta_{4,0}..., \beta_{4,z})$, $\boldsymbol{\gamma}$=$(\gamma_{1,0}..., \gamma_{1,s}, ..., \gamma_{4,0}..., \gamma_{4,z})$ and $\boldsymbol{\sigma}$=$(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, the set of MLE estimators $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\sigma}}\}$maximizes the log-likelihood function (15).

Estimation Results Table 2 shows the estimation results for the scale heterogeneous multinomial logit model. Table 3 shows the coefficients estimated after adjusting for scale heterogeneity among different treatment groups, with the significant levels for each coefficient inherited from Table 2. In Table 3, the key variables have expected signs and the coefficient estimates show a certain degree of consistency across different treatments (or the preference elicitation methods). We find that individuals in Treatment 3 (RC, using the incentive compatible preference elicitation method) are most sensitive to the cost changes and would exhibit the lowest willingness to pay. Individuals under the Treatment 1, Hypothetical referendum, are the least sensitive to the cost changes. This result is consistent with a large body of literature on the hypothetical bias in the stated preference experiment. The quantity coefficient and the co-benefit coefficients are positive and mostly significant, except individuals seem to not place a positive value on the co-benefit agricultural viability ($AgVia$) under the hypothetical referendum, even though the estimated coefficient is insignificant at a 10% level.

Besides comparing willingness to pay differences across preference elicitation methods, another way of examining the consistency of coefficients estimates is to see whether the marginal rate of substitution is constant between different types of co-benefits in each treatment. For example, the marginal rate of substitution between co-benefit z and z' in treatment t and t' under the null hypotheses $\frac{\partial U_t}{\partial z} / \frac{\partial U_t}{\partial z'} = \frac{\partial U_{t'}}{\partial z} / \frac{\partial U_{t'}}{\partial z'}$, which implies

[2] If individual i choose the buying nothing option N, the utility V_{in} is normalized to 0.

$\frac{\beta_{t,z}}{\beta_{t,z'}} = \frac{\beta_{t',z}}{\beta_{t',z'}}$. A weaker hypothesis is that the coefficient for each type of co-benefit is consistent across different treatments in terms of *relative* magnitude. Our results (Table 3) suggest the co-benefit $AgVia$ is the least preferred across all treatments. The Treatment HR, RC and RR_RC suggest the co-benefit $Habitat$ is the most preferred co-benefit, while the Treatment RR suggests $ReRunoff$ is the most preferred co-benefit. Therefore, though our estimates demonstrate comparable coefficient estimates results, we find the relative magnitude for different types of co-benefit do not always match with the results in Treatment RC, where each individual has incentive to always choose the utility maximizing option. Under the assumption that the incentive compatibility is necessary to estimate the "true" underlying utility function, the coefficients in other treatment are then biased due to the presence of potential strategic opportunities where not all individuals will choose the option that maximizes the utility among the set of alternatives, or simply, they do not have the economic incentive to think about the best option and choose accordingly. In the next section, we propose a method that explicitly accounts for such strategic opportunities or insufficient incentives and estimate the percentage of strategic choices in each treatment.

5 Non-Utility-Maximizing Choices and Valuation Implications

When a treatment is not incentive compatible, individuals may strategically or simply not choose the choice that yields the highest utility in a choice question. In our study, only Treatment 3, Real Choice with Incentive Compatibility (RC) is able to generate data that is consistent with the discrete choice model; that is, the observed choice maximizes utility among all three options in that specific choice experiment question. In the remaining of the paper, we use the phrase "utility-maximizing" option to refer the option that gives the highest utility among a given set of alternatives in *one* choice question based on the described attributes and cost. Note that an individual could still choose optimally by choosing the "non-utility-maximizing option" when she is behaving strategically to influence the group outcome through the choice of a second best or a third best option. Our Treatment 1 (HR), Treatment 2 (RR) and Treatment 4 (RR_RC) are considered not incentive compatible; the observed individual choice would not necessarily maximize utility among all three options due to 1) no economic incentive to choose the utility-maximizing option (e.g., Treatment 1) or 2) strategic incentives so that individual may choose option other than the utility-maximizing option when the aggregated group choices are used to influence the policy outcome (e.g., Treatment 2 or 4). Note that even though the Treatment 4 is not strictly incentive compatible, the gain from truthfully revealing one's preference may outweigh the cost (information rent) of not choosing the utility-maximizing option. In either case, we need to modify the discrete choice model to account for the preference misrepresentations and strategic behaviors when a treatment is not incentive compatible. Below we propose a model to estimate the percentage of truthful responses in each treatment when multiple treatments (including the incentive compatible treatment) are used in the experiment.

Still, we consider individual i's utility from choosing option j is:

$$U_{ij} = U(X_j, S_i, c_j) = V(X_j, S_i, c_j) + \epsilon_{ij}. \tag{16}$$

As noted before, in our experiment, each individual considers three alternatives: options A, B or the buying nothing option N. If the individual's choice implies his or her utility is higher for alternative $j \in \{A, B, N\} \equiv J$, providing utility U_{ij} compared to all the other alternatives $U_{ik}(k \neq j, k \in J)$, then the probability that individual i chooses alternative j is calculated by

$$\begin{aligned} P_i(j) &= Pr(U_{ij} > U_{ik}, k \neq j, k \in J) \\ &= Pr(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, k \neq j, k \in J) \\ &= Pr(\epsilon_{ij} - \epsilon_{ik} > V_{ik} - V_{ij}, k \neq j, k \in J), \end{aligned} \tag{17}$$

if individual i chooses the utility-maximizing option j, which only holds in the RC treatment. Thus, we assume all the observed choices are the utility-maximizing options in RC treatment and the proportion of utility-maximizing choices is $\pi_3 = 1$. The proportions of utility-maximizing choices in HR, RR and RR_RC are denoted as π_1, π_2 and π_4, respectively, with $\pi_t \in [0, 1]$.[3]

When $\pi_t \in [0, 1]$ $(t = 1, 2, 4)$, $1 - \pi_t$ of the observed choices do not maximizes the utility among the three options in a choice question. Thus, the true utility-maximizing choice is different than the observed one. Since there is no clear guidance on which of the remaining two options is the utility-maximizing option, we assume that one of the two remaining option has a probability of τ of being the utility-maximizing option while the other remaining option has a probability of $1 - \tau$ of being the utility-maximizing option. For example, if the option A is chosen from $\{A, B, N\} = J$, under our assumption, there is a probability π_t that $U_{iA} \in \max(U_{ik}), k \in J$, a probability $\tau(1 - \pi)$ (or $(1 - \tau)(1 - \pi)$)

[3]Note we assume that same underlying utility function and assume the difference between treatments are a result of the misrepresentation of the utility maximizing choices, the scale parameter does not apply in this situation.

that $U_{iB} \in \max(U_{ik}), k \in J$, or a probability $(1-\tau)(1-\pi)$ (or $\tau(1-\pi)$) that $U_{iN} \in \max(U_{ik}), k \in J$. Therefore, when option j is observed, the probability that individual i chooses this alternative is

$$\begin{aligned} P_i(j) &= \pi_t Pr(U_{ij} > U_{ik_1}, U_{ik_2}) + \tau(1-\pi_t)Pr(U_{ik_1} > U_{ik_2}, U_{ij}) + (1-\tau)(1-\pi_t)Pr(U_{ik_2} > U_{ik_1}, U_{ij}), \\ &= \pi_t \frac{e^{V_{ij}}}{1+\sum_{m=\{A,B\}} e^{V_{im}}} + \tau(1-\pi_t)\frac{e^{V_{ik_1}}}{1+\sum_{m=\{A,B\}} e^{V_{im}}} + (1-\tau)(1-\pi_t)\frac{e^{V_{ik_2}}}{1+\sum_{m=\{A,B\}} e^{V_{im}}}, \end{aligned} \tag{18}$$

together with equation (11), the log-likelihood function can be written as

$$\begin{aligned} \ln L &= \sum_i \left(d_{ij} \ln \left(\frac{\pi_t e^{V_{ij}} + \tau(1-\pi_t)e^{V_{ik_1}} + (1-\tau)(1-\pi_t)e^{V_{ik_2}}}{1+\sum_{m=\{A,B\}} e^{V_{im}}} \right) \right) \\ &+ \sum_i \left((1-d_{ij}) \ln \left(\frac{\pi_t + \tau(1-\pi_t)e^{V_{ik_1}} + (1-\tau)(1-\pi_t)e^{V_{ik_2}}}{1+\sum_{m=\{A,B\}} e^{V_{im}}} \right) \right). \end{aligned} \tag{19}$$

Let $\boldsymbol{\alpha}=\alpha$, $\boldsymbol{\beta}=(\beta_0..., \beta_z)$, $\boldsymbol{\gamma}=(\gamma_0..., \gamma_s)$ and $\boldsymbol{\pi}=(\pi_1, \pi_2, \pi_4)$, the set of MLE estimators $\{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\pi}}\}$maximizes the log-likelihood function (19). Note that the π_3 is constrained to 1 in the likelihood function since we assume that individuals will always choose the utility maximizing option in the Treatment RC.

One identification challenge is that when the utility maximizing option is not chosen, we could not match the probabilities $\tau(1-\pi_t)$ and $(1-\tau)(1-\pi_t)$ with the unobserved options k_1 and k_2. In the estimation process, we assume if $\tilde{V}_{ik_1} \geq \tilde{V}_{ik_2}$, then $\tau \geq 1-\tau$ in each iteration of optimization process, which implies that if the estimated utility from the option k_1 is higher than the option k_2, then the option k_1 has a higher probability being the utility maximizing choice compared to option k_2. We choose five different levels of τ ranging from equal probability ($\tau = 0.5$) to extremely unequal probability ($\tau = 1$) associated the unobserved choice with a higher estimated utility. Estimation results are shown in Table 4.

There are several interesting results from Table 4. First, we find that Treatment HR has the highest percentage of non-utility-maximizing choices and RR_RC has the lowest percentage of non-utility-maximizing choices. This result is stable in terms of relative magnitude regardless of choice on τ. Second, we find about 75% to 92% individual choices are utility-maximizing (in given the set of alternatives) in Treatment HR, 86% to 96% individual choices are utility-maximizing in Treatment RR and above 90% individuals choices are utility-maximizing in Treatment RR_RC, which implies the policy consequentiality introduced in Treatment RR and the payment incentive introduced in Treatment RR_RC encourage individuals to choose the utility maximizing choices in the experiment. Third, comparing the estimates π_1 and π_2, we find that adding the policy consequentiality (all individuals still get the same expected payment) increases the percentage utility maximizing choices by about 4% to 10%; comparing the estimates for π_2 and π_3, we find that adding the payment incentive (even though it is not considered incentive compatible) increase the percentage utility maximizing choices by about 3% to 5% relative to the total sample. Fourth, we find the relative magnitude of co-benefits are consistent with the coefficients estimated in Treatment *RC* and *RR_RC* (Table 3), only the preference for *Carbon* and *ReRunoff* are switched in the new model and the preference for these two types of co-benefits are very close in *RR_RC* (Table 3). We also find that the increase of τ tends to reduce the percentage of utility maximizing choices for all treatments.

Recall that we assume two reasons behind the non-utility-maximizing choices. Individual may not have the economic incentive (e.g., in a hypothetical treatment) since the outcome would not influence their utility in any direct ways. Also, individual may choose the option to maximize the potential outcome that is collectively determined by the group's choice, instead of the utility-maximizing option in a given choice scenario. Our model is unable to separate these two effects. Nonetheless, we could get a rough idea of their respective influences on truthfully choices by comparing the estimated percentage π across different treatment. In the next section, we discuss the implications of choosing the non-utility-maximizing choices on the expected utility of an individual making one decision.

Efficiency Implications To calculate the efficiency implications from individuals? strategic choices, we use the Treatment 3 (*RC*) as the baseline and assume it achieves 100% efficiency when using the utility function based on the incentive compatible data. Therefore, for any credit type k, individual i's utility of choosing the option j with (Q^*, c^*, Z^*) is

$$\hat{V}_i^* = Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) exp(D(S)). \tag{20}$$

However, individual i may not always choose the utility-maximizing option when the treatment is not incentive compatible. Our model assumes that individual i will only choose the utility maximizing option j with a probability $\hat{\pi}$ and choose another option, j' with a probability $1 - \hat{\pi}$. Therefore, when faced with the same choice scenarios, individual i's expected utility under the incentive-incompatible treatment is,

$$\hat{V}_i' = \hat{\pi} Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) exp(D(S)) + (1 - \hat{\pi}) Q_{j'}'^{\hat{\beta}} c_j'^{\hat{\alpha}} exp(B(Z')) exp(D(S)) \tag{21}$$

with $Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) \geq Q_{j'}'^{\hat{\beta}} c_j'^{\hat{\alpha}} exp(B(Z'))$. Therefore, it is easy to infer that $\hat{V}_i^* \geq \hat{V}_i'$, which implies when faced with the set of choices, the (expected) utility (based on the chosen option) is strictly higher in the incentive compatible treatment compared to treatment where strategic opportunity opportunities exist. The cost due to the presence of strategic behaviors can be calculated from

$$Q_j^{*\hat{\beta}}(c_j^* + C_s)^{\hat{\alpha}} exp(B(Z^*)) = \pi Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) + (1-\pi) Q_{j'}^{\prime\hat{\beta}} c_{j'}^{\prime\hat{\alpha}} exp(B(Z')). \tag{22}$$

Therefore, we can get

$$\begin{aligned} Q_j^{*\hat{\beta}}(c_j^* + C_s)^{\hat{\alpha}} exp(B(Z^*)) &\leq \pi Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) + (1-\pi) Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) \\ &= Q_j^{*\hat{\beta}} c_j^{*\hat{\alpha}} exp(B(Z^*)) \Rightarrow \\ C_s &\leq \sqrt[\hat{\alpha}]{\pi c_j^{*\hat{\alpha}}} - c_j^* \\ &= c_j^*(\sqrt[\hat{\alpha}]{\pi} - 1). \end{aligned} \tag{23}$$

The above equation implies an easy way to calculate the *upper bound* of the cost due to the presence of strategic opportunity (or inconsequentiality). The term $\frac{C_s}{c_j^*}$ represents the ratio of strategic cost relative to the transaction cost.[4] We define $M \equiv 1 - \frac{C_s}{c_j^*}$ to represent the achieved "efficiency" measure when the treatment is not incentive compatible, where the cost due to the presence of strategic opportunities is accounted and subtracted from a baseline efficiency level of 100%. Based on the coefficients estimated in Table 4, we calculated the achieved efficiency measure for Treatment 1, 2 and 4, which are compared to an efficiency measure of 100% in Treatment 3. Figure 2 shows the results, as well as the bootstrapped 95% confidence intervals of the achieved efficiency measures for each treatment, under different assumptions regarding the parameter τ. We find the efficiency measure M is highest in the Treatment RR_RC and lowest in Treatment HR; the efficiency measure also decreases as the parameter τ increases. Note that our results only represent the upper bound and our proposed measure is close to the true value when the strategic opportunity is small (i.e., individuals are more likely to choose the utility-maximizing option, τ is small).

6 Conclusion

Our study compares four different preference elicitation methods and investigates the robustness of estimation results assuming the incentive compatible treatment is able to produce unbiased utility parameters. We included a hypothetical treatment (Treatment HR), a policy consequential treatment (Treatment RR), an incentive compatible treatment Treatment IC), a policy and payment consequential treatment (Treatment RR_IC)

[4] In our model, the cost needed to buy Q units of water quality credits.

but lacks incentive compatibility. Our model shows the estimated coefficients exhibit a certain degree of consistency across all treatments and the Treatment *RR_RC* is the closest to the incentive compatible treatment *RC*. The EPRI's Ohio River Water Quality project provides a unique opportunity to establish consequentiality in a field context. Our research utilizes this opportunity and develops a choice experiment survey that presents realistic and consequential choice scenarios to solicit individuals' preference to water quality improvement and its associated co-benefits. We completed the credits transactions with EPRI and fulfilled our commitments in experiment.[5]

Criticisms toward the choice experiment usually include 1) individuals may not have sufficient incentive or 2) individuals may choose strategically so that they will not always choose the option that maximizes utility among the set of given alternatives, i.e., they are not revealing true preference. If the observed choices were not the utility-maximizing options in the choice set, the standard application of discrete choice model would produce biased coefficient estimates since one of the fundamental assumptions in the discrete choice model is violated. We assume that the observed choices always maximize utility in a given choice set in the discrete choice model, which is no longer true when individuals can strategically choose the second-best or other options.

To address this concern, we develop a model that is able to account for the situations when individuals may not always reveal their true preference (or ma not always choose the utility-maximizing options in a choice question). Our results suggest that about 75% to 92% individual choices are utility maximizing (given the set of alternatives) in the hypothetical treatment and above 90% individuals choices are utility maximizing in a treatment that is both policy and payment consequential however not incentive compatible. There are multiple ways to incorporate this result and explore its valuation implications. One way is to investigate the cost due to presence of strategic behaviors (or simply insufficient incentives). We proposal an efficiency measure to account for the strategic cost and the result shows that the Treatment *RR_IC* performs the best according to this measure.

We view this paper as a first step to explicitly model and identify strategic actions and recognize that when a preference elicitation method is *not* incentive compatible, the underlying data generating process violates the utility maximizing assumption in the discrete choice model. The nature of maximum likelihood estimation procedure may require a significant proportion of observations to be produced through an incentive compatible treatment in order to yield robust coefficient estimates. Though our estimation results perform well in our experimental data sample and consistent with previous literature on

[5]See Liu and Swallow (2016) for more information.

hypothetical bias and consequentiality, we hope to apply this method to other studies and see if it could be generalized to a variety of contexts.

Acknowledgement We thank Jessica Fox at the Electricity Power Research Institute (EPRI) for numerous valuable insights on the Ohio River Basin Trading Project and the assistance of her team in helping us access and understand projects producing credits and co-benefits. We thank Christopher Elphick, Timothy Vadas and Joglekar Deepak for help coordinating the experiments. We thank the participants at the USDA/EPA National Workshop on Water Quality Trading in Lincoln, NE. Joseph Lucibello provided efficient research assistance. This project is partially funded by USDA (U.S. Department of Agriculture)/NIFA (National Institute of Food and Agriculture)/AFRI (Agriculture and Food Research Initiative) grant # 2014-67023-21813 and grant #2016-67023-24748, the Zwick Center for Food and Resource Policy at the University of Connecticut (UConn) and the UConn Agricultural Experiment Station. This research is approved by Institutional Review Board (IRB) at UConn under the protocol #X15-049. All opinions and errors remain our own.

References

Adamowicz, Wiktor, Peter Boxall, Michael Williams, and Jordan Louviere, "Stated preference approaches for measuring passive use values: choice experiments and contingent valuation," *American journal of agricultural economics*, 1998, *80* (1), 64–75. 1

Bishop, Richard C and Thomas A Heberlein, "Measuring values of extramarket goods: Are indirect measures biased?," *American journal of agricultural economics*, 1979, pp. 926–930. 1

Boxall, Peter, Wiktor L Adamowicz, and Amanda Moon, "Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement*," *Australian Journal of Agricultural and Resource Economics*, 2009, *53* (4), 503–519. 1

Carson, Richard T and Theodore Groves, "Incentive and informational properties of preference questions," *Environmental and resource economics*, 2007, *37* (1), 181–210. 1

_ , _ , and John A List, "Consequentiality: A theoretical and experimental exploration of a single binary choice," *Journal of the Association of Environmental and Resource Economists*, 2014, *1* (1), 171–207. 1

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman, "Who is (more) rational?," Technical Report, National Bureau of Economic Research 2011. 1

Clarke, Edward H, "Multipart pricing of public goods," *Public choice*, 1971, *11* (1), 17–33. 1

Czajkowski, Mikołaj, Nick Hanley, and Jacob LaRiviere, "Controlling for the effects of information in a public goods discrete choice model," *Environmental and Resource Economics*, 2014, pp. 1–22. 4

Day, Brett, Ian J Bateman, Richard T Carson, Diane Dupont, Jordan J Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang, "Ordering effects and choice set awareness in repeat-response stated preference studies," *Journal of Environmental Economics and Management*, 2012, *63* (1), 73–91. 1

DeShazo, JR and German Fermo, "Designing choice sets for stated preference methods: the effects of complexity on choice consistency," *Journal of Environmental Economics and management*, 2002, *44* (1), 123–143. 1

Diamond, Peter A and Jerry A Hausman, "Contingent valuation: Is some number better than no number?," *The Journal of economic perspectives*, 1994, pp. 45–64. 1

Esponda, Ignacio and Emanuel Vespa, "Hypothetical thinking and information extraction in the laboratory," *American Economic Journal: Microeconomics*, 2014, *6* (4), 180–202. 1

Fiebig, Denzil G, Michael P Keane, Jordan Louviere, and Nada Wasi, "The generalized multinomial logit model: accounting for scale and coefficient heterogeneity," *Marketing Science*, 2010, *29* (3), 393–421. 4

Hanemann, W Michael, "Welfare evaluations in contingent valuation experiments with discrete responses," *American journal of agricultural economics*, 1984, *66* (3), 332–341. 1, 4

Harin, Alexander, "The random-lottery incentive system. Can p˜ 1 experiments deductions be correct?," in "16th conference on the Foundations of Utility and Risk" 2014. 1

Hausman, Jerry, "Contingent valuation: from dubious to hopeless," *The Journal of Economic Perspectives*, 2012, pp. 43–56. 1

Holmes, Thomas P and Kevin J Boyle, "Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions," *Land Economics*, 2005, *81* (1), 114–126. 1

Horowitz, John K, "The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods," *Economics Letters*, 2006, *93* (1), 6–11. 1

Kafle, Achyut, Stephen K Swallow, and Elizabeth C Smith, "Does Public Funding Affect Preferred Tradeoffs and Crowd-In or Crowd-Out Willingness to Pay? A Watershed Management Case," *Environmental and Resource Economics*, 2015, *60* (3), 471–495. 1

Karni, Edi and Zvi Safra, "" Preference reversal" and the observability of preferences by experimental methods," *Econometrica: Journal of the Econometric Society*, 1987, pp. 675–685. 1

Kawagoe, Toshiji and Toru Mori, "Can the pivotal mechanism induce truth-telling? An experimental study," *Public Choice*, 2001, *108* (3-4), 331–354. 1

Kawai, Kei and Yasutora Watanabe, "Inferring strategic voting," *The American Economic Review*, 2013, *103* (2), 624–662. 1

Liu, Pengfei and Stephen K. Swallow, "Integrating cobenefits produced with water quality BMPs into credits markets: Conceptualization and experimental illustration for EPRI's Ohio River Basin trading," *Water Resources Research*, 2016, pp. n/a–n/a. 3, 5

Louviere, Jordan J, David A Hensher, and Joffre D Swait, *Stated choice methods: analysis and applications*, Cambridge University Press, 2000. 1

McFadden, Daniel, *Conditional Logit Analysis of Qualitative Choice Behavior*, New York: New York: Academic Press, 4, 4

Meyerhoff, Jürgen, Malte Oehlmann, and Priska Weller, "The influence of design dimensions on stated choices in an environmental context," *Environmental and Resource Economics*, 2014, pp. 1–23. 1

Mitani, Yohei and Nicholas E Flores, "Hypothetical bias reconsidered: Payment and provision uncertainties in a threshold provision mechanism," *Environmental and Resource Economics*, 2014, *59* (3), 433–454. 1

Opaluch, James J, Stephen K Swallow, Thomas Weaver, Christopher W Wessells, and Dennis Wichelns, "Evaluating impacts from noxious facilities: including public preferences in current siting mechanisms," *Journal of environmental economics and management*, 1993, *24* (1), 41–59. 1

Swallow, Stephen K, James J Opaluch, and Thomas F Weaver, "Siting noxious facilities: an approach that integrates technical, economic, and political considerations," *Land Economics*, 1992, pp. 283–301. 1

_ , Thomas Weaver, James J Opaluch, and Thomas S Michelman, "Heterogeneous preferences and aggregation in environmental policy analysis: a landfill siting case," *American Journal of Agricultural Economics*, 1994, *76* (3), 431–443. 1

Vossler, Christian A and Mary F Evans, "Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality," *Journal of Environmental Economics and Management*, 2009, *58* (3), 338–345. 1

_ **, Maurice Doyon, and Daniel Rondeau**, "Truth in consequentiality: theory and field evidence on discrete choice experiments," *American Economic Journal: Microeconomics*, 2012, *4* (4), 145–171. 1, 2, 3

Zhang, Jing and Wiktor L Adamowicz, "Unraveling the choice format effect: A context-dependent random utility model," *Land Economics*, 2011, *87* (4), 730–743. 1

Table 1: Variable Descriptions

Variables	Description
Units	Number of water quality credit, measured in pound.
Nitrogen	Types of water quality credit, nitrogen=1 if it is nitrogen credit and =0 if it is phosphorus credit.
Cost	Individual private cost of purchasing a credit bundle, measured in dollars.
Co-Benefits	
Ag. Viability	Anything that increases farm productivity (reduces production costs and improves profits), resiliency to weather variability or improves livestock health. Binary variable equals 1 if present.
Carbon Sequestration and Soil Health	Carbon Sequestration: Holding more carbon in the soil to reduce greenhouse gas emissions. Soil Health: Improve soil quality and soil health by keeping soil covered, minimize disturbances, use cover crops and rotate crops to feed the soil. Binary variable equals 1 if present.
Habitat Enhancement and Pollinator Habitat	Habitat Enhancement: Add more diversity and select species of vegetation that benefit wildlife by providing food and cover. Pollinator Habitat: Providing food, habitat and cover to pollinators including honeybees, solitary bees and other pollinators (e.g. bats). Binary variable equals 1 if present.
Reduce Excessive Run-off	Protect the soil from intense rains (by keeping residue on the ground, using cover crops, using grassed waterways and filter strips). Binary variable equals 1 if present.
Reduce Animal Stress and Mortality	Reduce Animal Stress and Mortality. Binary variable equals 1 if present.

Table 2: Estimation Results: Heterogenous Scale Multinomial Logit

	Coeff. Est.	Std. Err.
HR (Hypothetical Referendum)		
ln(Cost)_HR	-0.6925***	0.0611
ln(Quantity)_HR	0.7935***	0.0636
Nitrogen_HR	0.0768	0.1429
AgVia_HR	-0.1820	0.2077
Carbon_HR	0.2328	0.1870
Habitat_HR	1.8649***	0.2176
ReRunoff_HR	0.5997***	0.1783
ReAniMor_HR	0.7591***	0.2475
Constant_HR	-0.1682	0.2010
Scale_HR	n/a	n/a
RR (Real Referendum)		
ln(Cost)_RR	-1.0861***	0.0681
ln(Quantity)_RR	1.2584***	0.0748
Nitrogen_RR	0.1351	0.1504
AgVia_RR	0.3403*	0.2228
Carbon_RR	0.6502***	0.2045
Habitat_RR	0.9237***	0.2110
ReRunoff_RR	1.5145***	0.1816
ReAniMor_RR	1.1959***	0.2610
Constant_RR	-0.3300*	0.2332
Scale_RR	0.9949***	0.0910
RC_IC (Real Choice with IC)		
ln(Cost)_RC	-1.1694***	0.0456
ln(Quantity)_RC	0.7539***	0.0496
Nitrogen_RC	0.1263	0.1092
AgVia_RC	0.2860**	0.1603
Carbon_RC	0.8532***	0.1479
Habitat_RC	1.3354***	0.1664
ReRunoff_RC	0.7764***	0.1369
ReAniMor_RC	1.1949***	0.1821
Constant_RC	0.7408***	0.1520
Scale_RC	0.6922***	0.0577
RR_RC (Real Referendum with Real Choice)		
ln(Cost)_RR_RC	-0.9973***	0.0409
ln(Quantity)_RR_RC	1.0741***	0.0448
Nitrogen_RR_RC	0.1178*	0.0910
AgVia_RR_RC	0.2397**	0.1355
Carbon_RR_RC	0.6654***	0.1239
Habitat_RR_RC	0.8659***	0.1304
ReRunoff_RR_RC	0.6720***	0.1119
ReAniMor_RR_RC	0.7570***	0.1529
Constant_RR_RC	-0.3721***	0.1388
Scale_RR_RC	0.6250***	0.0559
Demo. Control	Yes	
No. of Obs.	1237	
Log-likelihood	-832.39	
Degree of Freedom	47	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Estimation Results: Scale Parameter Adjusted

	HR	RR	RC	RR_RC
	Adj. Coeff. Est.	Adj. Coeff. Est.	Adj. Coeff. Est.	Adj. Coeff. Est.
ln(Cost)	-0.6925***	-1.0916***	-1.6895***	-1.5957***
ln(Quantity)	0.7935***	1.2648***	1.0892***	1.7186***
Nitrogen	0.0768	0.1357	0.1824	0.1884*
AgVia	-0.1820	0.3420*	0.4132**	0.3835**
Carbon	0.2328	0.6535***	1.2326***	1.0647***
Habitat	1.8649***	0.9284***	1.9294***	1.3856***
ReRunoff	0.5997***	1.5222***	1.1218***	1.0752***
ReAniMor	0.7591***	1.2019***	1.7263***	1.2112***
Constant	-0.1682	-0.3317*	1.0702***	-0.5953***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Coefficient signifiant levels inherited from Table 1.

Table 4: Estimation Results: Strategic Responses Adjusted

	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$	$\tau = 1.0$
	Coeff. Est. (Std. Err.)	Coeff. Est. (Std. Err.)	Coeff. Est. (Std. Err.)	Coeff. Est. (Std. Err.)	Coeff. Est. (Std. Err.)	Coeff. Est. (Std. Err.)
ln(Cost)	-1.3708*** (0.0405)	-1.4853*** (0.0442)	-1.6544*** (0.0492)	-1.7624*** (0.0523)	-1.8296*** (0.0544)	-1.8623*** (0.0556)
ln(Quantity)	1.3937*** (0.0365)	1.5048*** (0.0395)	1.6685*** (0.0438)	1.7757*** (0.0466)	1.8508*** (0.0484)	1.8954*** (0.0496)
Nitrogen	0.1469** (0.0808)	0.1472** (0.0851)	0.1446* (0.0923)	0.1433** (0.098)	0.1461* (0.1025)	0.15* (0.1058)
AgVia	0.2815*** (0.1186)	0.3013*** (0.1253)	0.3316*** (0.1368)	0.3486*** (0.1457)	0.3527** (0.153)	0.3494*** (0.1585)
Carbon	0.8769*** (0.1152)	0.9318*** (0.1263)	1.0517*** (0.144)	1.162*** (0.156)	1.2468*** (0.1636)	1.308*** (0.1681)
Habitat	1.5954*** (0.1255)	1.6446*** (0.1361)	1.728*** (0.1538)	1.7929*** (0.1669)	1.8436*** (0.1764)	1.8636*** (0.183)
ReRunoff	1.2268*** (0.1033)	1.3677*** (0.1113)	1.5953*** (0.1244)	1.7599*** (0.1336)	1.8798*** (0.1402)	1.9702*** (0.1452)
ReAniMor	1.3431*** (0.1375)	1.4533*** (0.1453)	1.6229*** (0.1574)	1.7401*** (0.1667)	1.8212*** (0.1741)	1.87*** (0.1797)
Cons	-0.073 (0.124)	-0.08 (0.1348)	-0.1086 (0.15)	-0.141 (0.1596)	-0.1518 (0.1658)	-0.1447 (0.1693)
$\pi_1(HR)$	0.9197*** (0.0363)	0.8869*** (0.0353)	0.8369*** (0.0351)	0.8014*** (0.0351)	0.7744*** (0.036)	0.7548*** (0.0359)
$\pi_2(RR)$	0.9597*** (0.0248)	0.9418*** (0.0243)	0.913*** (0.025)	0.8911*** (0.0257)	0.8737*** (0.0261)	0.8601*** (0.0269)
$\pi_4(RR\_RC)$	1.000*** (0.0001)	0.976*** (0.0351)	0.9469*** (0.0263)	0.9259*** (0.0266)	0.9094*** (0.0272)	0.8967*** (0.0284)
Demo. Control	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	1237	1237	1237	1237	1237	1237
Log-likelihood	857.369	-852.4310	-843.5590	-832.4150	-820.1470	-808.6540
DF	14	14	14	14	14	14

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: An Example of Choice Experiment Question.

	Option A	Option B	Option C
Nutrient Mitigated	Nitrogen	Phosphorous	None
Number of Water Quality Credits	20	15	None
Cost of Actions to You	$30	$40	$0
Agriculture Viability	Presence	None	None
Carbon Sequestration	None	Presence	None
Habitat Enhancement	Presence	Presence	None
Reduce Excessive run-off	None	Presence	None
Soil health and erosion	None	None	None
Pollinator Habitat	None	Presence	None
Reduce Animal Stress and Mortality	None	None	None
Your Choice			

Figure 2: Upper Bounds of Efficiency Measures in Each Treatment