# Open-ended format in contingent valuation: Additional evidence for the new hope

Ewa Zawojska
University of Warsaw, Faculty of Economic Sciences, Poland

Pierre-Alexandre Mahieu
University of Nantes, LEMNA, France

Patrice Guillotreau
University of Nantes, LEMNA, France

**Abstract**

Recent stated preference literature has identified theory-based conditions under which an open-ended value elicitation format is incentive compatible, in the sense of providing respondents with incentives to reveal their preferences truthfully (Vossler and Holladay, 2016). Our research verifies empirically the importance of the theory-based conditions in a field open-ended survey in which Canadian citizens express their willingness to pay (WTP) for a project of development of new underwater turbines for renewable energy production in Nova Scotia. We compare 8 model specifications, each assuming a different parametric distribution of WTP for the renewable energy project, and choose the one that fits best to the open-ended stated preference data. We find that WTP is on average lower and less noisy when it is derived from responses in a questionnaire that meets closer the incentive compatibility conditions. A formal statistical test reveals, however, that mean WTP estimates based on incentive-compatible and non-incentive-compatible questionnaires are not significantly different. Our evidence supports the findings of Vossler and Holladay (2016).

*Keywords*: stated preferences, contingent valuation, open-ended survey, incentive compatibility

*JEL classification*: C93; H41; Q42; Q51

## 1. Introduction

Stated preference surveys are widely used to determine values of public goods needed in cost-benefit policy assessments as well as in litigation processes. Although the use of stated preference methods dates back to early 1960s[1], serious doubts exist whether the methods reveal true values. In order to address this problem, researchers have been working on determining what survey design incentives respondents to disclose their preferences truthfully; a so-called incentive-compatible design. The conditions for incentive compatibility vary across different formats of value elicitation questions. A recent work of Vossler and Holladay (2016) identifies a set of theory-based conditions necessary for incentive compatibility of open-ended stated preference surveys. Our research aims at verifying the importance of these conditions in a field study.

Several general conditions need to be fulfilled for incentive compatibility of a stated preference survey (Carson and Groves, 2007; Vossler, Doyon and Rondeau, 2012). First, the policy project proposal presented in a survey must be viewed by a respondent as a take-it-or-leave-it offer, in the

[1] Ciriacy-Wantrup (1947) is thought as an initiator of development of stated preference methods, though Davis (1963) applied them for the first time in practice (Carson, 2001).

sense that a respondent's choice will not affect any other offers made in the survey. Second, the authority should be able to enforce the payment related to the project implementation; in other words, a payment vehicle should be coercive. Third, the survey needs to be seen by a respondent as consequential: a respondent should care about the project being valued and believe that each vote for the proposal at least weakly increases the chances of its implementation. Furthermore, a format of the value elicitation question plays a crucial role for incentive compatibility.

For long, a single binary choice question has been recommended for value elicitation in stated preference surveys because its incentive-compatible properties are well-known (Arrow et al., 1993). Despite this recommendation, researchers continue to use alternative value elicitation formats in order to gather broader information about preferences from a respondents' single answer. For example, a response to a multiple-option choice question or to an open-ended question will inform more about a respondent's preferences than a response to a binary choice question. Obviously, the broader information is obtained at the cost of possible bias of elicited values resulting from the format's lack of incentive compatibility. It does not imply, however, that formats other than a single binary choice question cannot be incentive compatible. They can, but maintaining their incentive compatibility imposes additional requirements on the survey design. For instance, if a sequence of binary choice questions is used, a researcher should assure that respondents assess each of the questions independently of other questions presented in the sequence (Vossler, Doyon and Rondeau, 2012).

Recently on the basis of theory, Vossler and Holladay (2016) identified conditions necessary for maintaining incentive compatibility of an open-ended stated preference survey. In addition to the general conditions listed above, an open-ended survey should inform respondents that the exact cost of implementing the proposed project is not known at the time of the survey and, therefore, respondents are asked to state their maximum willingness to pay (WTP) for it. The survey also needs to say that a respondent's stated WTP amount will be later treated as a vote in an advisory referendum; namely when the cost of the project becomes known, each declared WTP will be translated into a yes or no vote depending on whether the WTP is higher or lower, respectively, than the cost. The conditions require that respondents are aware that policy makers will use only the result of the advisory referendum, not the exact WTP amounts stated by respondents, for deciding whether to implement the project or not.[2]

Our research contributes to the existing literature twofold. First, we verify the role of the theory-based incentive compatibility conditions for an open-ended stated preference survey in a field valuation study. To the best of our knowledge, the research of Vossler and Holladay (2016) is the only available empirical inquiry that tests in the field the importance of the incentive compatibility conditions for open-ended surveys. We aim at providing additional evidence, in particular, evidence in a different valuation context. While the survey of Vossler and Holladay (2016) elicited preferences towards flood control system, we examine individuals' preferences towards renewable energy development. Second, we test how respondents' behaviour is affected by augmenting the survey script with explicit information that only the result of the advisory referendum, not the exact stated WTP amounts, will be communicated to policy makers. Vossler and Holladay (2016) found that WTP amounts reported in incentive-compatible and non-incentive-compatible questionnaires are not statistically significantly different. We extend their survey script by providing respondents with explicit information about what survey results will be later shared with policy makers and verify how this information impinges on respondents' behaviour.

[2] For intuition behind the mentioned conditions for incentive compatibility of an open-ended stated preference survey see Vossler and Holladay (2016).

The remaining of the paper is structured as follows. Section 2 describes the study design. Second 3 explains the econometric approach. Section 4 presents the results. Section 5 concludes.

## 2. Study design

### 2.1. Background

Canada has substantial renewable resources that can be used to produce energy: moving water, wind, biomass, sun, geothermal resources and ocean. A large share of the country current energy supply comes from renewable resources, and their use for energy production is still being developed. The ocean is probably the least exploited energy source[3], although it has great potential in the country surrounded by oceans (except for its southern border). A particularly promising site for the use of ocean energy is the Bay of Fundy in Nova Scotia that has the highest tides in the world.

The project presented in the survey involved constructing and testing two new underwater turbines in the Bay of Fundy in order to develop the use of ocean energy. Construction of the turbines was supposed to help investigate the productivity of such turbines and their impact on the environment. The project evaluated in the survey corresponded to actions planned and partially undertaken in Nova Scotia which involved building new underwater turbines. The fact that the ocean energy development in the Bay of Fundy has been being debated in Canada emphasised consequentiality of our survey.

### 2.2. Survey structure

The questionnaire began with general information about the survey topic, the organisations conducting the survey (which included the University of Quebec in Rimouski and the University of Nantes in France) and the possible use of survey responses for policy purposes as the survey results would be shared with policy makers.

After a few warm-up questions related to environmental protection and renewable energy, the proposed policy project was stepwise introduced. The description began with informing respondents about underwater turbines in general, explaining how they operated to produce energy, and discussing their strong and weak sites. This information referred to other renewable and non-renewable energy sources for comparison, such as wind and solar energy, coal and oil. Next, the details of the project proposal were provided. As mentioned earlier, the project involved building two new underwater turbines in the Bay of Fundy to develop the use of ocean energy, by testing the productivity of new turbines and their impact on the natural environment.

The payment, introduced upon the project implementation, was defined as an additional monthly fee added to the electricity bill of every household in Canada for the next year (12 months). In line with the recommendations by Vossler and Holladay (2016), the questionnaire emphasised that the exact amount of the fee would depend on the cost of implementing the project, which was unknown at the time of conducting the survey. Then, depending on a treatment, additional information was presented or not. The differences between the treatments are discussed in the next subsection.

Following a standard budget reminder, an open-ended valuation question was displayed, asking about the highest amount a respondent would agree to pay monthly for the next year to have the described project implemented. The exact formulation of the value elicitation question was as follows: “What is the highest amount which you would agree to pay monthly for the next year to

[3] Based on information of the Government of Canada at www.nrcan.gc.ca/energy/renewable-electricity/7295 (retrieved on April 11, 2017).

have the described program (constructing and testing two huge underwater turbines at the bay of Fundy) implemented?". Subsequently, a few follow-up questions were used to elicit respondents' consequentiality perceptions, among others, and to collect socio-demographic data.

### 2.3. Experimental treatments

The primary goal of our study is to examine whether the theory-based conditions for incentive compatibility of open-ended stated preference surveys matter empirically. To address this question, we introduced three treatments and each of them was assigned a bit different questionnaire. The three questionnaires differed only in the degree to which they met the theoretically suggested conditions. Every respondent was randomly assigned to one of the treatments.

The questionnaires used in the treatments differed only in the information provided before the value elicitation question. The baseline questionnaire resembled a typical open-ended stated preference survey with no additional information which is needed for satisfying the incentive compatibility conditions. We, henceforth, refer to this treatment as STANDARD, and we use data from this treatment as a reference point in our analysis.

The second treatment, labelled as treatment IC, included an additional script as suggested by Vossler and Holladay (2016) that is needed for the survey incentive compatibility. The script was placed right before the value elicitation question and mirrored the formulation used by Vossler and Holladay (2016). Namely, it said that the cost of the project realisation was uncertain at the time of the survey and it presented voting in the survey as participation in an advisory referendum: when the cost of the project became known, the open-ended stated values would be translated into yes or no votes on the project proposal.

The third treatment, labelled as treatment IC INFO, used an extended script in comparison with Vossler and Holladay (2016) in order to reinforce the compliance of the survey with the incentive compatibility conditions. As suggested by Vossler and Holladay (2016), for an open-ended survey to be incentive compatible, it is needed that respondents are guaranteed that only the result of the advisory referendum will be shared with policy makers upon the cost being known. In particular, respondents should be informed that policy makers will not be given the detailed data on stated WTP values for the project but only the percentage of yes and no votes on the proposal for a known cost of the project implementation.
Thus, in treatment IC INFO, we augmented the script used in treatment IC with a statement saying that we would not provide policy makers with information about the specific WTP amounts indicated by respondents, but instead we would provide them with the referendum outcome, that is, whether the population was in favour of the project or not at the given, known by then cost.

### 2.4. Survey administration

The survey was developed based on pretesting the questionnaire with individuals and on a pilot study conducted on 169 respondents. The final survey was conducted online among Canadian citizens by a professional public opinion research company in September 2016. In total, 1,253 respondents were surveyed and they were equally spread across the treatments. In the analysis provided in Section 4, we drop 14 observations because these respondents reported WTP comprising an incredibly large share of their income. Socio-demographic characteristics of the respondents dropped from the database, along with their stated WTP, are provided in Appendix A.

## 3. Econometric approach

We use the open-ended stated WTP amounts to find what parametric distribution describes WTP in the study sample best. We do not arbitrary choose a specific distribution of WTP, as there is no a priori or theory-driven information about the distribution of WTP. Instead, we try several common parametric distributions and select the specification which fits best to our data in terms of the Akaike and Bayesian Information Criteria (AIC and BIC, respectively).[4] To fit the parameters of each distribution, we calculate the value of the probability density function at stated WTP of every respondent. This gives the respondent's contribution to the likelihood function. By summing up each respondent's contributions and maximising the resulting function with respect to the distribution parameters, we are able to derive maximum-likelihood-method estimates of every distribution and then compare the values of their AIC and BIC.

Because WTP is expected to be non-negative and a substantial share of zero WTP responses is identified, we use a zero-inflated version of each parametric distribution: the respondents who declared that they were not willing to pay anything are cumulated in a spike-discontinuity of the WTP distribution at zero (Kriström, 1997). In addition, to verify the effect of consequentiality perceptions on respondents' behaviour in the survey, we incorporate as an explanatory variable of the model parameters a binary-coded variable indicating whether a respondent perceived a survey as consequential or not. This way, we allow for heterogeneous effects, for example, for mean or variance of the distribution, or for the probability of not being willing to pay at all (that is, WTP = 0).

Formally, we represent the probability of individual $i$'s WTP being equal to the stated amount as $PDF\left(WTP_i, \boldsymbol{\beta}_i\right)$, where $PDF$ denotes a cumulative distribution function of the selected WTP distribution and $\boldsymbol{\beta}_i$ is a vector of the distribution parameters (for example, mean and standard deviation for the normal distribution). By making $\boldsymbol{\beta}_i$ dependent on individual $i$'s characteristics, the parameter vector becomes individual-specific. In our study, we include only one individual-specific variable, namely a binary-coded variable equal to 1 for respondents perceiving the survey as consequential and 0 for respondents not convinced about the survey consequentiality.

Because $PDF\left(WTP_i, \boldsymbol{\beta}_i\right)$ expresses individual $i$'s contribution to the likelihood function, we represent the log-likelihood function for a sample of $N$ individuals as

$$\log L = \sum_{i=1}^{N} \log\left[ PDF\left(WTP_i, \boldsymbol{\beta}_i\right)\right].$$

Maximisation of the above equation generates estimates of the parameters underlying the WTP distribution.[5]

## 4. Results

This section provides the results of our empirical inquiry. First, we discuss socio-demographic characteristics of the three samples of respondents distinguished by treatment and verify whether the samples differ significantly in these characteristics. Second, we report consequentiality perceptions across the samples and test their correlation with the treatments. Finally, we present the results of estimation of the models as described above.

[4] We compare fit of the models to data mainly based on AIC and BIC, instead of using the log-likelihood values, because the considered distributions differ in the number of parameters.

[5] The models were estimated using a custom code developed in Matlab, available at https://github.com/czaj/DistFit under CC BY 4.0 license.

### 4.1. Socio-demographic characteristics of treatment samples

Socio-demographic characteristics of the three samples distinguished by treatment are given in Table 1. In general, the samples do not differ substantially with respect to these characteristics. In each treatment sample, about a half of the respondents is female, has an academic degree and is married, while about one third of the respondents is single. An average respondent is about 48 years old and has income above 3,700 Canadian dollars (CAD).

Table 1. Socio-demographic characteristics of the treatment samples

| | **STANDARD** | **IC** | **IC INFO** |
|---|---|---|---|
| Number of respondents | 414 | 412 | 413 |
| Female | 49.03% | 48.79% | 51.82% |
| Single | 33.82% | 33.25% | 33.66% |
| Married | 56.52% | 56.31% | 54.96% |
| Divorced / Widowed | 9.66% | 10.44% | 11.38% |
| Academic degree | 44.69% | 41.50% | 41.89% |
| Age | 48.77 (17.11) | 48.48 (16.94) | 48.17 (17.28) |
| Income [CAD] | 3,763 (2,255) | 3,707 (2,243) | 3,723 (2,264) |

*Note*: Except for age and income, the table provides shares of respondents within each treatment sample. For age and income, means are reported, along with standard deviations in the brackets.

We formally test for differences in socio-demographic characteristics of the treatment samples. Table 2 presents the results of these tests. For the discrete variables, that is, for gender, marital status and academic degree, chi-squared tests of equality of proportions are conducted. For the continuous variables, that is, for age and income, Wilcoxon signed-rank tests are used, because the variables are not normally distributed. In the case of every comparison, we cannot reject the null hypothesis of equality of distributions, which implies that the treatment samples do not differ statistically significantly with respect to socio-demographic characteristics.

Table 2. Results of tests of differences in socio-demographic characteristics across the treatment samples

| | **STANDARD vs. IC** | | **STANDARD vs. IC INFO** | | **IC vs. IC INFO** | |
|---|---|---|---|---|---|---|
| | **Test statistic** | **P-value** | **Test statistic** | **P-value** | **Test statistic** | **P-value** |
| Female | $\chi^2 = 0.005$ | 0.943 | $\chi^2 = 0.640$ | 0.424 | $\chi^2 = 0.757$ | 0.384 |
| Marital status | $\chi^2 = 1.190$ | 0.755 | $\chi^2 = 4.557$ | 0.207 | $\chi^2 = 1.198$ | 0.753 |
| Academic degree | $\chi^2 = 0.852$ | 0.356 | $\chi^2 = 0.659$ | 0.417 | $\chi^2 = 0.012$ | 0.911 |
| Age | W = 86,436 | 0.737 | W = 87,350 | 0.589 | W = 85,756 | 0.843 |
| Income | W = 86,602 | 0.697 | W = 86,478 | 0.771 | W = 84,842 | 0.945 |

### 4.2. Consequentiality perceptions across treatments

Respondents' consequentiality perceptions are assessed with the use of two statements included in the survey. Namely respondents indicated the degree to which they agreed with each of the following sentences: (1) "the results of this survey will be used by policy makers in order to take the final decision regarding the program implementation" and (2) "I care about whether the program presented in this survey will be introduced". Assessment of consequentiality perceptions based on the two statements corresponds to the definition of consequentiality proposed by Carson and Groves (2007). Carson and Groves discerned two elements of consequentiality: "a survey's results are seen by the agent as potentially influencing an agency's actions" and "the agent cares about the outcomes of those actions" (p. 183).

Each of the two statements measuring consequentiality perceptions was evaluated on a Likert scale with five levels, which were labelled, respectively, as "I do not agree at all", "I do not agree", "I somewhat agree", "I agree" and "I fully agree". As the extreme categories were chosen less

frequently than the middle ones, we combine the first two and the last two levels. In Table 3, for every treatment sample, we report shares of the respondents in each category of consequentiality perceptions.

Table 3. Consequentiality perceptions by treatment (percentage shares of the respondents)

| | I do not believe in the survey's policy consequences. | | I somewhat believe in the survey's policy consequences. | | I believe in the survey's policy consequences. | |
|---|---|---|---|---|---|---|
| I do not care about the project introduction. | STANDARD: | 6.76 | STANDARD: | 1.93 | STANDARD: | 1.45 |
| | IC: | 7.28 | IC: | 2.43 | IC: | 2.43 |
| | IC INFO: | 9.69 | IC INFO: | 3.15 | IC INFO: | 2.66 |
| I somewhat care about the project introduction. | STANDARD: | 8.21 | STANDARD: | 17.87 | STANDARD: | 7.00 |
| | IC: | 8.50 | IC: | 15.53 | IC: | 5.83 |
| | IC INFO: | 7.51 | IC INFO: | 16.22 | IC INFO: | 8.23 |
| I care about the project introduction. | STANDARD: | 9.18 | STANDARD: | 17.63 | STANDARD: | 29.95 |
| | IC: | 9.47 | IC: | 16.02 | IC: | 32.52 |
| | IC INFO: | 6.30 | IC INFO: | 16.46 | IC INFO: | 29.78 |

*Note*: The shares sum to 100% for each treatment separately.

As shown in Table 3, the distribution of the respondents across the categories of consequentiality perceptions is very similar for every treatment. We observe that a large share of respondents was convinced about survey's consequentiality measured by the two dimensions as suggested by Carson and Groves (2007).

We formally verify with the use of chi-squared tests whether respondents differ in self-reported consequentiality perceptions across the treatments. Table 4 summarises the tests' results. For none of the comparisons, the null hypothesis of equality of distributions can be rejected, implying that the treatment samples do not differ in regard to consequentiality perceptions. This finding contradicts the observation by Vossler and Holladay (2016) who reported that standard and incentive-compatible open-ended formats gave rise to different distributions of consequentiality perceptions.

Table 4. Results of tests of differences in consequentiality perceptions across the treatment samples

| | **STANDARD vs. IC** | | **STANDARD vs. IC INFO** | | **IC vs. IC INFO** | |
|---|---|---|---|---|---|---|
| | **Test statistic** | **P-value** | **Test statistic** | **P-value** | **Test statistic** | **P-value** |
| Degree of agreement with: "I care about whether the program presented in this survey will be introduced" | 1.501 | 0.827 | 5.924 | 0.205 | 3.643 | 0.456 |
| Degree of agreement with: "The results of this survey will be used by policy makers in order to take the final decision regarding the program implementation" | 2.771 | 0.597 | 4.914 | 0.296 | 1.420 | 0.841 |

*Note*: For the purpose of testing, we use the respondents' original declarations of their degree of agreement with each consequentiality statement, that is, without recoding the responses into broader categories as done in Table 3.

For the need of further analysis and for the ease of interpretation, we combine the consequentiality perceptions self-reported by respondents in the two statements into a single binary-coded variable taking a value of 1 for those believing in the survey consequentiality and 0 for others. Namely we treat as consequential those respondents who:

- care about the project introduction and believe in the survey's policy consequences;
- somewhat care about the project introduction and believe in the survey's policy consequences;

- care about the project introduction and somewhat believe in the survey's policy consequences.
We do not include other categories into the group of consequential respondents, since their responses indicate that they might have some serious doubts about the survey's consequentiality. We use the binary-coded variable for the remainder of the paper.

We verify whether socio-demographic characteristics explain variability in consequentiality perceptions. We run a logistic regression with a dependent variable being the binary-coded variable equal to 1 for consequential respondents and 0 for others.[6] The results of the estimation are presented in Appendix B. The only variable that significantly explains consequentiality perceptions is age: the probability of viewing the survey as consequential decreases with age. The results also confirm that consequentiality perceptions are not related to treatments.

### 4.3. Differences in stated preferences across incentive-compatible and non-incentive-compatible surveys

Treatments differ in the degree of complying with the incentive compatibility conditions suggested for open-ended valuation surveys. Thus, a comparison of WTP across the treatments will reveal how the survey's adherence to incentive-compatible design impinges on respondent's behaviour.

For every treatment, we estimated 8 models, each assuming a different underlying distribution of WTP.[7] As explained in Section 3, every model includes a spike-discontinuity of the WTP distribution at zero and a binary-coded variable indicating whether a respondent believes in survey consequentiality or not.

Table 5 provides the summary of our estimation results. For every model, we report the values of a log-likelihood function (LL), AIC and BIC, where the two latter are corrected by the number of the respondents in a given treatment (n). The models listed in Table 5 are ordered according to the values of the information criteria from lowest to highest. In the case of every treatment, the order is identical, with the only exception of the extreme value and Rayleigh distributions being switched for treatment IC.

Table 5. Comparison of parametric distributions fitted to the open-ended data

| | STANDARD | | | IC | | | IC INFO | | |
|---|---|---|---|---|---|---|---|---|---|
| **Distribution** | **LL** | **AIC/n** | **BIC/n** | **LL** | **AIC/n** | **BIC/n** | **LL** | **AIC/n** | **BIC/n** |
| Log-normal | -1423.77 | 6.907 | 6.966 | -1422.02 | 6.932 | 6.991 | -1362.97 | 6.629 | 6.688 |
| Log-logistic | -1427.61 | 6.926 | 6.984 | -1424.31 | 6.943 | 7.002 | -1363.20 | 6.631 | 6.689 |
| Birnbaum-Saunders | -1430.66 | 6.940 | 6.999 | -1433.91 | 6.990 | 7.048 | -1375.97 | 6.692 | 6.751 |
| Exponential | -1544.89 | 7.483 | 7.522 | -1503.32 | 7.317 | 7.356 | -1431.17 | 6.950 | 6.989 |
| Logistic | -1790.73 | 8.680 | 8.738 | -1735.17 | 8.452 | 8.511 | -1649.85 | 8.019 | 8.077 |
| Normal | -1914.69 | 9.279 | 9.337 | -1857.47 | 9.046 | 9.105 | -1754.84 | 8.527 | 8.586 |
| Extreme value | -2013.31 | 9.755 | 9.814 | -2117.06 | 10.306 | 10.365 | -1911.82 | 9.287 | 9.346 |
| Rayleigh | -2093.81 | 10.134 | 10.173 | -1997.70 | 9.717 | 9.756 | -1923.48 | 9.334 | 9.373 |

[6] The findings are consistent if we take as a dependent variable respondents' original declarations to either of the two statements used for the assessment of consequentiality perceptions.

[7] The code in Matlab mentioned in footnote 5 and used in this study includes more distributions than the 8 presented here. They are, however, not considered in the current analysis because they do not match the distribution of WTP responses as shown by computational estimation difficulties related to those distributions.

The log-normal distribution fits the data best as measured by AIC and BIC. We provide the parameter estimates for the log-normal distribution for each treatment sample in Table 6.[8] The estimates take similar values across the samples, which suggests that the treatments might not affect WTP stated by respondents. Below we further investigate this issue and formally test for differences. The results in Table 6 also show that the spike constant is significant for every treatment. Furthermore, significantly different distribution parameters are observed for consequential respondents. In particular, the WTP distributions for those perceiving the survey as consequential in each treatment are characterised by high means and low levels of the spike constant when compared to non-consequential respondents.

Table 6. Estimation results of WTP for the log-normal distribution

| | **STANDARD** | | **IC** | | **IC INFO** | |
|---|---|---|---|---|---|---|
| | Mean parameter estimates | Parameter estimates for consequential (interaction) | Mean parameter estimates | Parameter estimates for consequential (interaction) | Mean parameter estimates | Parameter estimates for consequential (interaction) |
| **Location parameter** | 1.9732*** | 0.3647** | 1.9575*** | 0.4313*** | 1.7851*** | 0.5366*** |
| | (0.1446) | (0.1814) | (0.1321) | (0.1621) | (0.0730) | (0.0880) |
| **Scale parameter** | 1.5300*** | -0.0486 | 1.3851*** | -0.0654 | 1.3456*** | -0.1101 |
| | (0.1022) | (0.1282) | (0.0934) | (0.1146) | (0.0924) | (0.1114) |
| **Spike constant** | -0.2423*** | -0.6346*** | -0.2150** | -0.9573*** | -0.1607* | -0.9924*** |
| | (0.0924) | (0.1333) | (0.0922) | (0.1423) | (0.0918) | (0.1412) |

*Notes*: *** and ** denote significance at the 1% and 5% levels, respectively.

In order to verify differences in WTP between the treatments and between the consequential and non-consequential respondents, we use the bootstrapping technique proposed by Krinsky and Robb (1986) to simulate the parameters of the WTP distribution. The simulation results are provided in Table 7. Several important findings emerge from this evidence showing how WTP is affected by treatments and by consequentiality perceptions.

Table 7. Simulated parameters of the log-normal distribution of WTP

| | **Mean (μ)** | **Standard deviation (σ)** | **Coefficient of variation (σ/ μ)** | **95% confidence interval for the mean** | **Spike probability** |
|---|---|---|---|---|---|
| **STANDARD** | | | | | |
| Consequential | 25.4202 | 68.8694 | 2.7092 | [18.3411; 38.3324] | 0.1924 |
| Non-consequential | 13.8463 | 46.2145 | 3.3377 | [8.9126; 23.2262] | 0.4057 |
| **IC** | | | | | |
| Consequential | 22.8904 | 49.6832 | 2.1705 | [17.1994; 30.7145] | 0.1216 |
| Non-consequential | 10.8355 | 30.8163 | 2.8440 | [7.3323; 16.7305] | 0.4148 |
| **IC INFO** | | | | | |
| Consequential | 18.8293 | 36.1677 | 1.9208 | [14.7486; 24.6952] | 0.1254 |
| Non-consequential | 8.1606 | 22.9639 | 2.8140 | [5.6058; 12.4740] | 0.4365 |

Focusing first on the effect of treatments, we observe that the mean WTP values are highest in treatment STANDARD and lowest in treatment IC INFO. This suggests that WTP gets on average lower as the survey approaches an incentive-compatible design. However, a comparison of corresponding 95% confidence intervals reveals that mean WTP values do not differ significantly across the treatments. Although Vossler and Holladay (2016) reported higher mean WTP for an incentive-compatible open-ended survey than for a standard open-ended survey, they also found no statistically significant differences in WTP between the two surveys.

[8] Our findings reported in the remainder of this section are robust to other specifications of the WTP distribution such as log-logistic, Birnbaum-Saunders and exponential.

A clear patter also emerges in terms of changes in standard deviations by treatment: as we move from a survey adhering less to a survey adhering more to the incentive compatibility conditions, standard deviations of the WTP estimates become smaller. This finding is confirmed when the standard deviations are corrected for the differences in means, namely the coefficients of variation are smaller for a survey that meets incentive compatibility conditions closer. Our findings overlap with those of Vossler and Holladay (2016), who reported more noisy value estimates for a standard open-ended survey in comparison with an incentive-compatible open-ended survey.

We now move to the discussion of the effect of consequentiality perceptions on stated WTP. Within each treatment, respondents believing in the survey consequentiality appear to declare on average substantially higher WTP values than respondents not convinced about the survey consequentiality. Smaller coefficients of variation for the former than for the latter imply that WTP values stated by consequential respondents are less dispersed. For those believing in the survey being consequential, the spike probabilities take considerably lower levels, which corresponds to their substantially higher mean WTP. Our finding of higher WTP values for consequential respondents is consistent with the results reported by Vossler and Holladay (2016).

### 4.4. Respondents' perceptions on the survey's adherence with incentive-compatibility conditions and differences in stated preferences across respondents with different perceptions

Even if the survey script contains necessary information to satisfy incentive compatibility conditions, the survey may be not incentive compatible in the views of respondents. For example, empirical evidence shows that including additional scripts informing about the survey's consequential nature does not affect respondents' perceptions about the survey's consequentiality (Czajkowski et al., 2016, Drichoutis et al., 2015, Oehlmann and Meyerhoff, 2016). Thus, in this subsection, we focus on preferences of the respondents who not only participated in the most incentive-compatible survey (treatment IC INFO), but who also saw the survey as satisfying the conditions for incentive compatibility. We inquire whether their preferences differ from the rest of the treatment sample.
In order to assess respondents' perceptions on the survey's adherence to incentive-compatibility conditions, we included at the end of a questionnaire several follow-up statements and asked respondents for the degree of their agreement with each of them. Two of the statements, as already explained in subsection 4.2, help measure respondents' beliefs in the survey's consequentiality. Additionally, we include two other statements to verify whether respondents believe that the cost of implementation of the proposed project is unknown at the time of the survey and that only the percentage share of yes and no votes will be provided to policy makers. The exact formulations of the statements corresponding to the two beliefs are as follows: "the exact cost of the program implementation is unknown yet" and "only the outcome of the referendum (not specific responses provided by respondents) will be shared with policy makers". Agreement with each statement was indicated on a five-degree Likert scale with the same levels as used by consequentiality statements, that is, "I do not agree at all", "I do not agree", "I somewhat agree", "I agree" and "I fully agree".

In order to obtain WTP of respondents who are presented with an incentive-compatible survey and at the same time see the survey as adhering to the incentive compatibility conditions, we choose from the sample in treatment IC INFO such respondents who agreed or fully agreed with the statements about the cost of the project being unknown and the referendum outcome being shared with policy makers. The simulated parameters of the log-normal distribution of WTP for these respondents are as presented in Table 8.

Table 8. Simulated parameters of the log-normal distribution of WTP for incentive-compatible respondents

| | |
|---|---|
| **Mean (μ)** | 25.6505 |
| **Standard deviation (σ)** | 50.9569 |
| **Coefficient of variation (σ/ μ)** | 1.9866 |
| **95% confidence interval for the mean** | [18.3589; 37.4535] |
| **Spike probability** | 0.1074 |

When the preferences of the respondents perceiving that the survey meets the incentive compatibility conditions are compared with the preferences of the consequential respondents in this treatment (treatment IC INFO reported in Table 7), we observe that WTP of the former is on average substantially higher and a bit noisier. However, the 95% confidence intervals for the mean overlap for the two groups of respondents implying that the difference in mean WTP is not statistically significant.

## 5. Conclusions

Recently, Vossler and Holladay (2016) identified a set of theory-based conditions needed for incentive compatibility of open-ended stated preference surveys. In a field study of preferences of Canadian citizens towards development of renewable energy, we verify how satisfying the conditions specific to open-ended surveys impinges on respondents' behaviour. Similarly to Vossler and Holladay (2016), we find that for a survey that meets the incentive compatibility conditions closer, WTP is on average lower and less noisy.

**References**

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register*, 58(10), 4601-4614.

Carson, R. T., and Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1), 181-210.

Czajkowski, M., Vossler, C. A., Budziński, W., Wiśniewska, A., and Zawojska, E. (2015). Addressing empirical challenges related to the incentive compatibility of stated preference methods. Working paper no. 31(179), Faculty of Economic Sciences, University of Warsaw.

Drichoutis, A. C., Vassilopoulos, A., Lusk, J. L., and Nayga Jr, R. M. (2015). Reference dependence, consequentiality and social desirability in value elicitation: A study of fair labor labeling. 143rd Joint EAAE/AAEA Seminar, Naples, Italy, March 25-27.

Kriström, B. (1997). Spike models in contingent valuation. *American Journal of Agricultural Economics*, 79(3), 1013-1023.

Krinsky, I., and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 715-719.

Oehlmann, M., and Meyerhoff, J. (2016). Stated preferences towards renewable energy alternatives in Germany – Do the consequentiality of the survey and trust in institutions matter? Journal of Environmental Economics and Policy, 6(1), 1-16.

Vossler, C. A., Doyon, M., and Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, 4(4), 145-171.

Vossler, C. A., and Holladay, J. (2016). *Alternative value elicitation formats in contingent valuation: A New Hope*. Working paper no. 2016-02, University of Tennessee, Department of Economics.

**Appendix A**

Characteristics of the respondents dropped from the study sample

| | Treatment | WTP | Gender | Marital status | Attained education | Age | Income [CAD] |
|---|---|---|---|---|---|---|---|
| 1 | STANDARD | 250 | Male | Single | No academic degree | 53 | Below 1,000 |
| 2 | STANDARD | 2,222 | Male | Single | PhD or higher | 39 | 2,000-3,000 |
| 3 | STANDARD | 1,000 | Female | Married | No academic degree | 73 | 3,000-4,000 |
| 4 | IC | 5,000 | Male | Married | PhD or higher | 44 | 5,000-6,000 |
| 5 | IC | 50,000 | Male | Single | No academic degree | 24 | Above 7,000 |
| 6 | IC | 1,500 | Female | Married | PhD or higher | 28 | 4,000-5,000 |
| 7 | IC | 1,000 | Female | Single | No academic degree | 21 | 1,000-2,000 |
| 8 | IC | 9,000,000 | Male | Single | Master's | 22 | Below 1,000 |
| 9 | IC INFO | 450 | Female | Single | Master's | 21 | 1,000-2,000 |
| 10 | IC INFO | 10,000 | Female | Married | PhD or higher | 36 | 5,000-6,000 |
| 11 | IC INFO | 1,000 | Male | Married | No academic degree | 61 | 1,000-2,000 |
| 12 | IC INFO | 2,000 | Female | Divorced | Master's | 43 | 2,000-3,000 |
| 13 | IC INFO | 700 | Female | Single | Master's | 22 | 1,000-2,000 |
| 14 | IC INFO | 100,000 | Female | Married | No academic degree | 18 | 2,000-3,000 |

**Appendix B**

Results of a logistic regression of consequentiality perceptions

| Variable | Estimate | Standard Error |
|---|---|---|
| Female | -0.206 | 0.118 |
| Married | -0.059 | 0.124 |
| Academic degree | 0.071 | 0.119 |
| Age | -0.014*** | 0.004 |
| Income | 0.000 | 0.000 |
| IC | -0.009 | 0.141 |
| IC INFO | -0.004 | 0.141 |
| Constant | 0.785*** | 0.229 |
| **Model diagnostics** | | |
| Log-likelihood | -843.50 | |
| AIC | 1703.00 | |
| BIC | 1743.98 | |

*Notes*: The dependent variable is a binary-coded variable equal to 1 for respondents believing in the survey consequentiality and 0 for others. *** denotes significance at the 1% level.