Sample size and consequentiality in stated preferences

Abstract

Sample size has received little attention in the growing literature on consequentiality. An online single-bounded dichotomous choice field study dealing with underwater turbines, a part of participants is provided with information on the number of persons that will participate to the survey, where the information varies across participants. Our main result is that the information on sample size has no effect on willingness-to-pay, which suggests that consequential binary contingent valuation studies can still ensure incentive-compatible behavior when an expected gain from voting becomes very small.

1 Introduction

In the political literature, the concept of pivotally correspond to the probability of a given vote to change the outcome of the election. Intuitively, a rational individual will go voting if the benefits from voting are superior to the cost. Some empirical evidence gives support to this rational voter model. For instance, De Paola and Scoppa (2014) shows that people are more likely to go voting when the degree of competition between candidates is high; while Martins (1995) shows that the turnout in municipal election is lower in big cities than in low cities.

There have been important concerns about the validity of the stated preference surveys over the last decades. Among these concerns, it has been argued that the best strategy for rationale participants when voting is not necessarily to respond truthfully. To improve the validity of stated preference surveys, Carson and Groves (2007) proposed some conditions for truthful preference revelation. These conditions include the use of a single binary choice (SBC) survey format and also consequentiality, which means respondents believe that their choices in a survey might have consequences in real life. In a following-up work, Vossler et al. (2012) formalized the different conditions for a SBC to be incentive compatible. Among them, the participants should believe that a "yes" or "no" response would increase or decrease the probability for the program to be implemented. For instance, if the sample size drastically increases, the probability for a given response to be "pivotal" (i.e., to change the decision of the policy maker) will decrease but it will not reach zero, hence the survey will remain incentive compatible.

Mitani and Flores (2012) considered the case where responding to a valuation task involves an effort[1] and raises the following question: "Do consequential binary CV referenda still ensure incentive-compatible behavior when an expected gain from voting becomes very small?" If the cost from voting is superior to the benefit, it is unclear why the rational individual would involve an effort in the valuation task and respond truthfully. As the sample size increases, it may no longer be worth investing an effort in the valuation task. The authors realized an induced value experiment and vary, among other things, sample size. The authors found that varying the sample size from 1 participant to 45 participants had no impact on results (in their case the proportion of error). To the best of our knowledge, the impact of

[1] The authors state: "The cost of voting means any cost caused by making a voting decision, including a cognitive task of judging which alternative is better, time to make a vote, and/or participation in the voting decision".

sample size on stated preference studies has never been tested in a field study, where the number of participants is generally substantially higher than 45 participants.

In this paper, we report the results of an online split sample survey dealing with a water turbines program in France that was conducted in March 2018 by a professional company who guaranteed in advance a number of participants. A part of the sample is not provided with information on the sample size, like in most surveys. Another part is provided information just before the valuation question where the information varies across participants, with one of the three following sentences being randomly allocated to each of the participant: "at least 50 participants will participate to the survey", "at least 200 participants will participate to the survey" and "at least 2000 participants will participate to the survey". The sentence was constructed in a way to vary the "perceived sample size" while avoiding deception.

We find that the information on sample size does not affect WTP, which is consistent with the study from Mitanes and Flores (2012), nor does it affect the response time which can reflect the effort invested in the valuation task. Furthermore, we find that people who think that the survey can influence policy makers (generally referred to as "policy consequentiality" in the literature) has a positive effect on WTP, which is consistent with past studies. A growing number of studies assess the level of policy consequentiality on a rating type question, like "do you think your response and the one from the other respondent will influence policy makers". This type of question may capture/grasp two different effects simultaneously: a) to which extend people think that their response can influence the outcome of the study, which is the main focus of our paper, and will be called hereafter as "response consequentiality", and/or b) to which extend they think the outcome of the study can influence policy makers, which we will refer to as "survey consequentiality". Our result suggests that the "response consequentiality" does not play a key role in the effect of "policy consequentiality" on WTP.

The remaining of the paper is structured as follows. Section 2 describes the survey. Section 3 presents the results. Section 4 provides a discussion and Section 5 concludes.

2 Survey

In France, a large part of the electricity comes from nuclear power (about 75%) and the share is expected to decrease to 50% by 2025 according to the French energy transition for green growth act voted in 2015[2]. Several programs of off shore wind turbines and underwater turbines are being considered, although in March 2018, when our survey was performed, none of the planned off shore and underwater turbines were operating. The program described in the survey was based on an existing project of constructing big underwater turbines to improve the productivity of this type of renewable energy.

The questionnaire was structured as follows. First, it was explained that the responses will be communicated to policy makers and therefore might influence policy makers. Second, information was provided on renewable energy and more specifically on underwater turbines. Pros and cons of underwater turbines were explained. Third, participants were described with a research program consisting of constructing and setting up two giant underwater turbines, of 16 meters in France. It was explained that the effect of the underwater turbines on fauna and flora would be studied and that the two turbines would produce electricity for about 5,000 households. The location of the underwater turbines was not provided. Fourth, the following

[2] https://www.ecologique-solidaire.gouv.fr/loi-transition-energetique-croissance-verte

valuation task was asked: “Would you be willing to pay X EUR a month during a year on your electricity bill for the set-up of this program (building and testing underwater turbines)?” Voluntary payment was avoided to avoid free riding. The following final bid amounts were retained based on a pre-test: 0.5; 2; 5; 10 and 20 EUR. Fifth, different debriefing questions were asked regarding the perception of the program and socio-demographic questions. Among other things, rating type questions were asked, with people being asked to assess how they agreed with the following statements: “This project is very important for France” and “The outcome of this survey will influence policy makers on the decision to implement or not the program”. The latter statement aims at measuring “survey consequentiality”, which is a component of “policy consequentiality”.

The survey was performed by a professional company which guaranteed that they would deliver a representative sample of the French population of at least 2,000 participants (the actual number was 2,023). The representativeness of sample was based on three variables: gender, income and age. Regarding the experimental design, each individual was randomly allocated to one of four versions of the questionnaire. In the baseline version of the questionnaire, hereafter called V1, no information was provided on the questionnaire. In the other versions, one sentence was added before the valuation question: “at least 50 participants will participate to the survey” in V2, “at least 200 participants will participate to the survey” in V3 and “at least 2000 participants will participate to the survey” in V4. To ensure that people would read the sentence, there was little information on the slide in addition to the sentence (see Appendix A).

Focus groups and pre-tests showed that the survey was properly worded and that some persons were in favor of the program because it could potentially harm fauna and flora. In the final surveys, some of the person refused to pay for the program for this reason. Hence, possible negative WTP should be accounted for in the econometric treatment.

3 Results

Table 1 provides information on socio-demographic characteristics of the respondents, where the continuous variable income correspond to net monthly income (expressed in thousands of euro) and the binary variable education take value 1 if the individual has at least the A-level. A non-parametric Kolmogorov test is performed successively for each of the six possible combinations (V1 versus V2; V2 versus V3, etc…) and for each of the four socio-demographic variables (income, age, education, female). Results indicate that there is no difference of distribution across sub-samples at 5% statistical level for each of the four socio-demographic variables, which is not surprising given that the allocation to the different versions of the questionnaire was random.

Table 1 Sociodemographic characteristics

	V1 (no info) n=468	V2 (>50) n=532	V3 (>200) n=520	V4 (>2000) n=503
Income	2.702 (1.36)	2.750 (1.523)	2.633 (1.435)	2.677 (1.475)
Age	46.111 (15.102)	47.130 (15.587)	46.238 (14.994)	45.386 (15.763)

Education	0.765 (0.424)	0.759 (0.428)	0.738 (0.44)	0.750 (0.434)
Female	0.506 (0.500)	0.504 (0.500)	0.498 (0.500)	0.513 (0.500)

In the rest of the paper, we consider several ways to construct the variables related to sample size, as can be seen in Table 3. This allows testing if the results are sensitive to the variable construction. The variable "sample size information" involves the all the sub-samples (V1, V2, V3 and V4), while the other variables only involves three sub-samples (V2, V3 and V4) since the baseline (V1) is excluded.

Table 2 Construction of variables related to sample size information

Versions 1, 2, 3 and 4 of the questionnaire	
"Sample size information"	Binary variable that takes value 1 if the respondent has been assigned the questionnaire version 1 (V1, no info), 0 if version 2 (V2 >50), version 3 (V3 >200) or 4 (V3 >2000)
Version 2, 3 or 4 of the questionnaire	
Sample size	Continuous variable that takes value 50 if the respondent is assigned version 2 of the questionnaire (V2 >50), 200 if version 3 (V3 >200) and 2,000 if version 4 (V4 >2000).
Sample_200	Binary variable that takes value 1 if the respondent has been assigned the version 3 (V3 >200) of the questionnaire, zero if version 2 (V2 >50) or 4 (V4 >2000).
Sample_2000	Binary variable that takes value 1 if the respondent has been assigned version 4 (V4 >2000) of the questionnaire, zero if version 2 (V2 >50) or 4 (V3 >200)

The interval data regression model (Cameron, 1988) is employed to explore the determinants of WTP. This approach relies on the maximum likelihood estimation approach, which requires to assume a distribution for WTP. We retain the normal distribution to account for the possible negative WTP due to the possible effect of the program of fauna and flora. Therefore, WTP_i is a linear function of a row vector of covariates, x_i such that $\text{WTP}_i = x_i\beta + \varepsilon_i$ where β is a column vector of unknown parameters and ε_i is a normally distributed mean-zero error term with standard deviation σ_i. Since the variance of the error term may depend on the experimental design (i.e., the variance may differ across sample size), we allow for heteroscasticity (see Vossler and Hollaway, 2018 and Vossler and Zawojska, 2018 for recent examples using the same approach). In this interval data model, "yes" and "no" responses are considered as censored data, since the only information which is observed is whether the WTP is above or below the assigned bid amount.

Table 3 Interval data regression model

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Coefficient parameters						
Income				1.259***	1.506***	1.539***
				(0.372)	(0.445)	(0.452)
Education				2.014	1.134	1.242
				(1.243)	(1.480)	(1.510)
Female				-3.803***	-4.150***	-4.133***
				(1.028)	(1.240)	(1.263)
Age				0.00137	0.0192	0.0161
				(0.0340)	(0.0408)	(0.0412)
Important				14.98***	15.20***	15.24***
				(1.913)	(2.270)	(2.304)
Survey consequentiality				6.859***	7.975***	8.163***
				(1.325)	(1.629)	(1.660)
Sample size info	0.636			1.054		
	(0.994)			(1.119)		
Info_200		-1.288			-1.277	
		(1.330)			(1.386)	
Info_200		-0.816			-0.764	
		(1.509)			(1.555)	
Log(sample)			-0.170			-0.170
			(0.372)			(0.399)
Constant	6.696***	8.073***	8.269***	-14.81***	-14.73***	-14.70***
	(0.821)	(1.044)	(2.118)	(3.205)	(3.768)	(4.349)
Standard errors						
Sample size info	0.235			0.0941		
	(0.149)			(0.120)		
Info_200		-0.255			-0.155	
		(0.195)			(0.140)	
Info_2000		0.0257			0.0931	
		(0.219)			(0.154)	
Log(sample)			0.0170			0.0292
			(0.0523)			(0.0387)
Constant	2.603***	2.920***	2.743***	2.727***	2.874***	2.698***
	(0.124)	(0.151)	(0.299)	(0.117)	(0.117)	(0.228)
Observations	2,023	1,555	1,555	2,023	1,555	1,555

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 shows that the information on sample size has no effect on willingness-to-pay, regardless of the sample size variable construction. The coefficients are not statistically significant from zero at conventional levels. The same results are obtained when including socio-demographic and follow-up questions. Results suggest that the variance of the error term is also not statistically significant at a conventional statistical level, hence suggesting that increasing the sample size does not lead to more random answers.

We also find that the level of income has a positive effect on WTP, which is consistent with a priori expectation. Also, people who believe that "the outcome of this survey will influence policy makers on the decision to implement or not the program" tend to state higher willingness-to-pay, where the variable (called "survey consequentiality") takes value 1 if the respondents report a score of at least 3 on the range going from 1 ("I do not agree at all") to 5 ("I fully agree"). Based on the same threshold (3), people who believe that the "program is important for France" tend to report a higher WTP.

We also created a series of interaction variables between the variables related to sample size (e.g., sample size info, info_200,..) and the variable on sociodemographic questions (e.g., education, age, etc), but none of the interaction variables turned-out to be statistically significant at conventional levels. We also investigated the combined effect of sample size and debriefing questions. For instance, we created an interaction variable between the sample size and the "survey consequentiality" to check if the combination of both components of "policy consequentiality" ("response consequentiality" and "survey consequentiality") could have an impact. However, none of the interaction variables turn out to be statistically significant. We also express the WTP in logarithm form, hence assuming a lognormal distribution for WTP, but again similar results were obtained as with a normal distribution[3].

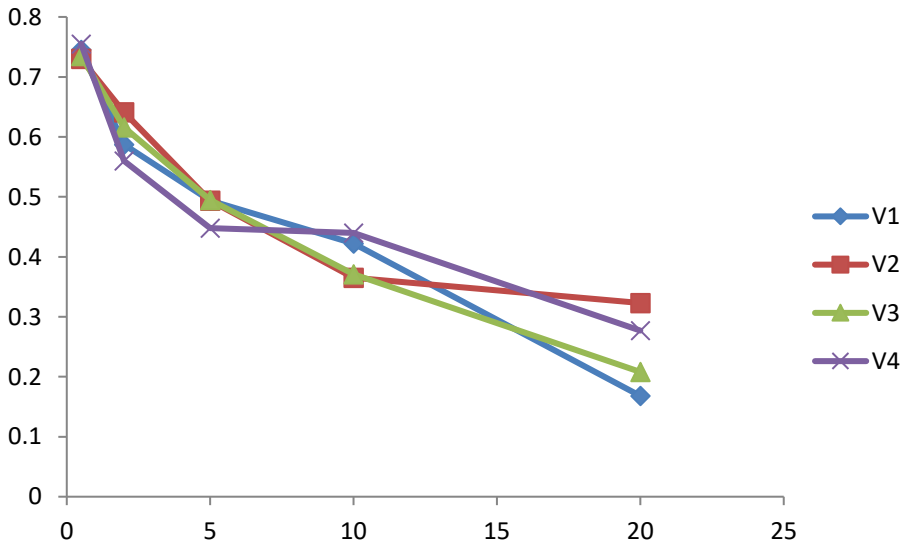
Table 4 displays the mean WTP and confidence interval from the interval data model without covariate and the Turnbull approach. The null hypothesis of equal mean WTP is rejected for each of the possible combinations and for both parametric and non-parametric analysis. Figure 3 shows the survival distribution for each of the treatments. The survival distributions are close, hence supporting the finding that the information on sample size has no effect on willingness to pay.

[3] Results are available upon request.

Table 4. Parametric and non-parametric mean WTP and confidence intervals

	Description	V1 (Ø)	V2 (>50)	V3 (>200)	V4 (>2000)
Mean WTP (Std error)					
Parametric	Mean WTP	6.696 (0.821)	8.073 (1.044)	6.785 (0.825)	7.257 (1.090)
	Confidence interval	[5.086; 8.304]	[6.027; 10.119]	[5.169;8.401]	[5.121; 9.394]
Non-parametric	Turnbull Mean WTP	6.525 (0.497)	7.864 (0.548)	6.712 (0.490)	7.532 (0.543)
	Confidence interval	[5.551;7.499]	[6.790;8.938]	[5.752 ;7,672]	[6.468;8,596]

Figure 1. Survival distribution



We now turn to the analysis of response time, where the response time corresponds to the number of seconds taken by participants to answer the following valuation question: “Would you be willing to pay X EUR a month during a year on your electricity bill for the set-up of this program (building and testing underwater turbines)?” A full slide was devoted to the valuation question, with no additional script/information being displayed on it except the “yes” or “no” answer. Table 5 shows that there is little difference across treatments. For instance, the median is identical (11 seconds). The non-parametric Kolmogorov test is employed to compare the distributions and the null hypothesis of identical distribution is never rejected.

Table 5 Number of seconds taken to answer the valuation question

Percentile	V1	V2	V3	V4
1%	4	4	4	3
5%	5	6	5	5
10%	7	6	6	6
25%	8	8	8	8
50%	11	11	11	11
75%	14	14	15	14
90%	18	21	22	20

4 Discussion

We find that the information on sample size has no effect on people's behavior, which is consistent with the results found by Mitani and Flores (2012) on an induced experiment that involved 45 participants and suggest that responses provided by respondents in a consequential SDC surveys are pretty robust, which gives some support to the use of SDC. This result also suggests that even if people had a priori expectation on sample size in SDC surveys, this is not necessarily a problem as sample size does not impact willingness-to-pay or response time. Another contribution of our paper regarding the literature on consequentiality is to highlight that "survey consequentiality" is an important component of the "policy consequentiality". Indeed, the variable related to "survey consequentiality" is statistically significant in our interval data model, unlike the variables related to sample size which refers to the "response consequentiality". If so, practioners should keep highlighting in the questionnaire that the outcome of survey will be communicated to policy makers, as recommended in several guidelines (e.g., Johnston et al. 2017).

Our study suffers from possible limitations. First, the information on sample size can affect protest answers or create a selection bias (i.e., people refusing to take part of the survey if the sample is too high or too low), and therefore alter the comparison across treatments. However, we did not provide the information on sample size at the beginning of the survey to avoid selection bias. As for the protest answers, the rate of protest answers is not different across treatments and removing protest answers do not change any of the results[4] Another limitation is that we cannot guarantee that people understand the link between pivotally and sample size. Some people might not fully understand that a larger sample makes each response less influential. However, results from the political literature suggest that the potential number of voters matters. Finally, providing information on sample size can change the perception of the good. When the sample size is big, participants may think that the good to be valued is

[4] The results of the statistical treatment without the protest answers is available upon request

"important". If so, difference of mean willingness to pay would have been observed across treatments. In addition, we do not find any correlation between the variable "interest" and the sample size, hence suggesting that it does not change the perception of the good.

5 Conclusion

In this paper, we test in a single-bounded dichotomous choice survey dealing with wind turbines whether providing information on sample size has an impact on willingness-to-pay and respond time. We find that there is no effect, and conclude that consequential binary contingent valuation studies can still ensure incentive-compatible behavior when an expected gain from voting becomes very small. Overall, this gives some support to the use of the single bounded dichotomous choice surveys. Future studies could check if similar results can be found when the format involves a higher cognitive burden, such as choice experiment. Conditional voting requires an important effort (i.e., predict the vote for each of the alternatives and voting among the top two alternatives) and it might not be worth investing this effort when the sample size is high and the probability to affect the outcome is therefore low.

Appendix A Slide positioned just before the valuation question

Version of the questionnaire: V1 (no information)

- The results of this survey will be communicated to policy makers

Version of the questionnaire: V2 (>50)

- At least 50 persons will participate to the survey
- The results of this survey will be communicated to policy makers

Version of the questionnaire: V3 (>200)

- At least 200 persons will participate to the survey
- The results of this survey will be communicated to policy makers

Version 4 (information on sample size, >2000)

- At least 2000 persons will participate to the survey
- The results of this survey will be communicated to policy makers

References

Cameron T.A., 1988. A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. Journal of environmental economics and management 15, 355-379

Carson T.R. and T. Groves, 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics. 37, 181–210

De Paola M. and V. Scoppa, 2014.The impact of closeness on electoral participation exploiting the Italian double ballot system. Public Choice,160, 467–479.

Johnston R.J., Boyle K.J., Adamowicz W., Bennett J., Brouwer R., Cameron T. A., Hanemann W. M., Hanley N., Ryan M., Scarpa R., Tourangeau R. and Vossler C. A., 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists 4, no. 2, 319–405.

Martins M R, 1995. Size of Municipalities, Efficiency, and Citizen Participation: A Cross-European Perspective. Environment and Planning C: Politics and Space, vol. 13, 441-458

Mitani Y. and N. E. Flores. 2012. Robustness Tests of Incentive Compatible Referenda: Consequential Probability, Group Size, and Value-cost Difference. Paper presented at the 19th Annual Conference of European Association of Environmental and Resource Economists, June 27−30, Prague, Czech Republic.

Vossler C.A., M. Doyon and D. Rondeau, 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics 4(4): 145-171

Vossler C.A. and J. S. Holladay, 2018. Alternative value elicitation formats in contingent valuation: Mechanism design and convergent validity. Journal of Public Economics 165: 133-145.

Vossler, Christian A. and Ewa Zawojska. 2018. Toward a better understanding of elicitation effects in stated preference studies. Working Paper #2018-01, Department of Economics, University of Tennessee.