

# **Sampling and selection bias in internet panels used for Contingent Valuation surveys – an empirical investigation concerning an environmental non-market good**

Ole Bonnichsen<sup>1</sup> and Søren Bøye Olsen<sup>1</sup>

*<sup>1</sup>Environmental and Natural Resource Economics Unit, Institute of Food and Resource Economics, Faculty of Life Sciences, University of Copenhagen, Denmark*

In the context of economic valuation of non-market goods based on stated preferences, data collection by means of internet surveys have gained widespread use in recent years. The most severe disadvantage of using internet surveys is associated with the sampling procedure. As access to a computer and the internet is not typically available to every single individual in a population, problems concerning sample coverage and sample representativeness may be expected. While a lack of internet penetration may not seem as large a concern today as previously, there is still a proportion of any population who never use the internet. This constitutes a coverage problem as this group of people will be excluded from entering an internet survey of the general public. Another drawback of using the internet includes the uncertainty of whether or not there might be additional selection biases introduced. People can choose whether or not to be part of an internet panel and secondly also whether they wish to participate in the survey, thereby introducing two levels of potential selection bias. The decision to be part of an internet panel and subsequently respond to a survey may be correlated with people's preferences thus making the respondents a non-random and non-representative sample with regard to preferences in the population, ultimately biasing results from the survey (Heckman 1979).

Therefore, in order to generalize valuation estimates from a sample of respondents in a stated preference survey to the intended target population, the panel and subsequent sample should reflect the target population in the tested area with regard to socioeconomic characteristics as well as overall preference structures of the population. If this is not the case then it would not be possible to make valid inferences and extrapolate value estimates from the sample to the target population. Included in this paper is an analysis investigating the representativeness of a sample with regards to central socioeconomic characteristics. This analysis is done on three levels from 1) the target population of the tested area in Denmark to 2) the internet panel population and finally 3) the respondent population (i.e. the sample). With this procedure it will be possible to ascertain if estimates of Willingness-To-Pay (WTP) could potentially be biased due to the sample not representing the target population.

The data used for the analysis originates from an internet Contingent Valuation survey eliciting preferences for improvements in water quality of a river in Denmark. The internet panel company sent the survey out to a total of 2000 individuals who were asked a screening question of whether or not they wanted to participate in the survey (henceforth in the paper named ANSWER and NO\_ANSWER respectively). The NO\_ANSWER respondents had been previously asked a series

of questions regarding their socioeconomic characteristics upon recruitment into the panel, thereby facilitating a comparison. Data regarding the characteristics of the target population were obtained from Statistics Denmark. It should be noted that, since we have so far only done preliminary analyses, results and tests mentioned in the following are subject to change and primarily reflect the structure of analysis we intend to report in the final paper.

The results of the socioeconomic characteristics comparison between the three levels of participation indicate that there are significant differences between the target population, the panel population, and the final sample. This suggests that indeed selection bias could be a problem in this case.

Since WTP is elicited using a payment card approach in the survey, the stated maximum WTP values are censored within the amount interval presented in the payment card. Hence, we use a Maximum Likelihood interval estimation method for a Grouped data model which is considered a variant of the censored regression model (Cameron and Huppert, 1989) in order to find significant determinants of WTP. Furthermore, we ran probit models to check for significant differences between the ANSWER and NO\_ANSWER respondents to identify the factors that affect the probability of answering. The dependent variable takes the value 1 if the respondent has answered and zero if the respondent has not answered. We find that the factors that affect the decision of respondents to participate in the internet panel and in the survey (respectively) differ from those that affect their answers to the valuation question. Hence, respondents essentially face a two-stage decision process. Firstly, they decide whether they will participate at all and then secondly they state their maximum WTP.

We additionally find that some of the variables that affect the decision to participate also affect the WTP. This would suggest that those who choose not to participate could have a different WTP than those who choose to participate and state a WTP (i.e. WTP estimates are likely to suffer from selection bias). We then proceeded with an extension of the grouped data model that takes the selection bias into account and corrects for it. Since the Heckman two-step procedure is only appropriate for a continuous dependent variable, we estimate this model using Full Information Maximum Likelihood. Here the significance of the correlation parameter between the participation and valuation equation in the model is tested to see if a selection bias is present between the populations and the sample. If it is below zero, the mean WTP would be underestimated whereas the opposite would be the case if it is above zero. However, if the correlation is not significantly different from zero then the two decision processes can be considered independent, suggesting that the concern about potential selection bias can be relaxed. Our preliminary results suggest that the correlation parameter is significantly positive, suggesting that a selection bias is present in our data which, unless incorporated as in our model, causes overestimation of the WTP.

## **References**

- Cameron, T.A. and Huppert, D.D. (1989): OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management* 17(3): 230-246.
- Heckman, J.J. (1979): Sample selection bias as a specification error. *Econometrica* 47(1): 153-161.