

Dear Prof. Cameron:

Thank you for your manuscript, "Comprehensive selectivity assessment for a major consumer panel: attitudes toward government regulation of environment, health and safety risks" (POQ-05-0336), recently submitted to the journal.

Based on the reviewer comments, which can be found at the bottom of this letter, and my own reading of the paper, I regret to inform you that we cannot publish your manuscript. The journal is able to publish only one of every ten submissions. As you can see from the reviews that follow, one of the reviewers was quite positive about the manuscript but the other two reviewers expressed serious concerns.

I appreciate your interest in Public Opinion Quarterly and thank you for giving us the opportunity to review your research. I hope the comments enclosed will be useful to you in your future work in this area.

Sincerely,

Nancy Mathiowetz  
Associate Editor, Public Opinion Quarterly  
nancym2@uwm.edu

Reviewer(s)' Comments to Author:

**Reviewer: 1**

Comments to the Author

Referee Report: Comprehensive Selectivity Assessment for a Major Consumer Panel: Attitudes toward Government Regulation of Environment, Health and Safety Risks

I fear I missed the point of this paper. The goal is to determine whether responses to a question on governments role in regulating environment, health and safety hazards contained within a private large scale consumer survey. You apply textbook selection procedures and find little evidence that self selection by the respondents bias the response.

Several questions remain unanswered. Why do we expect there to be a bias on nonresponse? Is there empirical analysis supporting the conjectured relationship of ones political views and participation in the survey? Without evidence supporting the possible contamination of selection bias the paper reads as if it is attacking a straw man.

Second, and related to the first, why is unit nonresponse modeled versus item non-response? Item non-response strikes me as more the more natural forum. Besides by analyzing item non-response, responses for the current survey may serve as possible instruments. And of course the factors that affect unit nonresponse may be different than governing item non-response.

Perhaps this was explained in the redacted text, but I wondered how the online and webTV survey verified the identity of the respondent. You use census tract information to proxy characteristics of the household, how then do you know who in the household answers the survey? How is the identity of the respondent confirmed to validate survey responses? And as a panel, if you trust the survey responses, then I dont see the need for the census tract. You can use responses from prior rounds to estimate the first stage probit.

Only a single sentence lists the target question. Yet we are not told where in the survey it is placed. One might expect different responses after a sequence of questions on taxes paid, than say after questions on use of national and state parks. Contextual issues in the framing of the target question strike me as at least as large as those from self-selection.

It would have been useful to see if selection appears in the earlier stage samples. I appreciate the work that was done constructing the samples and the attempt to get a representative sample. It would be interesting to see if less carefully constructed samples exhibit more self selection bias.

The statistical analysis strikes me as flawed. As you note the outcome variable is ordinal. Consequently, you should estimate a bivariate order probit and not a Heckman selection. Your analysis imposes an arbitrary cardinality that is not innocuous in the analysis of the possibility of selection. It is not that much more difficult to estimate the order-probit at the second stage as the full information maximum likelihood Heckman correction.

Doesnt the selection problem go away if the analysis is done at the census tract level and not at

the individual household level? This is tantamount to assuming the same response pattern across census tracts, conditional on the latent factors. What unobserved component is thus left to bias the relationship?

On a minor point, you note (line 5-6, page 18) that standard errors in models 3 and 4 are inaccurate. A number of textbooks (e.g., Woolridge) present straightforward discussions of their correction. The paper is so careful on some aspects of the statistical analysis and inattentive on others.

Finally, let me close with a quibble. On page 8, lines 10-20 you claim that your factor score estimates are population values. I agree with 523506 observations sampling error should be negligible and disregarded (though I wonder about the need to estimate regressions with 523 odd thousand observations and why we test at conventional 5 and 10 percent levels), but disagree you have recovered population values. You have recovered values at a point in time but the population values may change over time. It is a small point and I agree it doesn't affect your statistical inference.

## **Reviewer: 2**

### Comments to the Author

This is a very thorough and competent analysis of sample selection issues in a survey sample source of growing importance. The research methods and findings are relevant to a wide variety of researchers and government agencies. There are some limitations to the reliance of zip code data at the initial stages of the selection models, but the authors acknowledge these and handle this in an appropriate manner. The paper is generally very well written.

### Key comments:

1. The modeling approach in section 4 needs references or further justification.
2. Some editing is suggested for section 4.
3. For reasons given below, consider dropping models 3 and 5 and summarizing the minimal differences in a footnote.

### More specific edits and comments:

Page 3: line 8, change though to through

Page 5, line 6: the factor analysis is unavailable for review, but is likely sound based on quality signals I can glean from the paper

Page 5, line 9: Would subset 2 have about 7000 cases? Why not indicate that here?

Page 6, footnote 4: why not state the number of contacts that were lost (i.e., how many is a very limited set...?)

Page 7, footnote 5: For a tiny minority of Census tracts again, why not state the number?

Page 9, first four sentences of section 3 inclusive of footnote 8: This part is unnecessary for this paper. Moreover, it distracts readers. I presume that it refers to a stated choice survey with many versions based on some orthogonal experiential design, but the readers of this paper likely don't care. Regarding footnote 8, introducing the more complicated methods that could have been used in another paper is out of place in this paper.

Page 11. First full sentence. Change observed to something like observed and in some cases simulated. Otherwise it is confusing since the sample doesn't change and observed data doesn't change, so why does the distribution and mean change? This becomes more (somewhat) clear later in the paper, but is inaccurate as currently stated.

Page 14, footnote 11: I'm still a little unsure about the procedure for predicting/simulating the mean and distribution of the selectivity corrected dependent variable. When it says add back in the error term does this really mean add back in the residual? Or, does it mean that errors were repeatedly drawn to simulate the distribution? Please clarify.

Page 15: Section 4. The rationale for 4 is a little rough, lacks citations, and is quite apologetic.

Page 18, first paragraph. I suggest dropping the first sentence, and then cutting the remaining part of the paragraph and moving it up so that it ends the second paragraph of section 4 (top of page 16).

Page 15-19, section 4: I am not aware of any other papers that follow this exact approach to diagnosing selection biases. Can the authors provide any references for similar efforts to model selection by using deviations from the mean selection index (prob)? Intuitively, it makes sense to me, and surely as the selection bias goes to zero the models become the standard OLS, but who else has done this?

Page 16, line 4: change result to results

Page 16, line 6: change an to any

Page 17, entire second paragraph and its footnote: This material should appear in an earlier section of the paper I suggest at the end of section 2.1 on page 6. (The paper discussed in the footnote looks like an interesting one!)

I would drop models 3 and 5. Why? First, to me, the prob indexed models are more intuitive because the index has a very straightforward interpretation. Second, models 3 and 4 are not at all different so nothing is lost here. Third, models 5 and 6 have very minor differences that ultimately dont much affect the model. Thus, it makes sense to streamline the tables and discussion by dropping models 3 and 5. The minor differences between model 5 and 6 can be mentioned in a footnote and the reader will trust you (as was done for the ordered probit models).

Page 19, line 5: change employs to employ

Page 19, first paragraph: See my question above for footnote 11. I like this comparison, but please clarify.

Page 21, line 15: change typically a more to typically more

Thats it.

### **Reviewer: 3**

#### Comments to the Author

This manuscript attempts to specify the factors that contribute to non-representativeness in a consumer panel that is maintained by a market research firm. Overall, the manuscript represents an ambitious attempt to grapple with complex statistical issues that are likely to surface when analysts construct random samples. Given the complexity of the econometric models that appear in the paper, I have no doubt that the authors are very competent researchers. Notwithstanding my praise for the technical expertise of the authors, for two reasons, I cannot endorse publication of the manuscript in its present form. First, given that the overall goal of the paper is to assess the presence of sample selection bias that results from observed and unobserved characteristics, the authors do little to explain any of the technical concepts that undergird their analyses, leaving the reader to infer the meaning of technical concepts from the authors' allusions to them. Second, the authors do not come to any firm substantive conclusions in their work other than the statistical results that come out of the very scholarly analyses. I elaborate further on these topics below.

#### Definition and Explanation of Technical Concepts

In the authors' introduction, they argue that the non-representativeness of a sample can result if the sample and the population differ in terms of unmeasured or unobserved characteristics. In essence, it appears to me that the majority of this piece is about modeling the effects of these unmeasured factors and how they contribute to sample selection bias, as evidenced by their research questions that appear on page 3. However, the authors fail to fully define what is meant by sample selection bias specifically. If econometric novices are interested in making sure that the samples they draw via RDD techniques are representative of the population, but are unaware of the technical aspects of dealing with unmeasured or unobservable factors, then a thorough definition of sample selection bias is needed. But the authors do not provide one. This is a serious flaw because an understanding of the problem of sample selection bias is critical to understanding why the authors use the methods they do and why they compare these methods to standard techniques (more on this point below).

The authors' assumption of intricate econometric knowledge of readers is manifested by their use of sample selection bias as a conceptual framework for the paper and selectivity-corrected models (page 3) as statistical methods for demonstrating the strength of their arguments. For example, there is an entire section dedicated to an apparent definition of the Heckman Selectivity Correction (page 10-12). However, nowhere do the authors specify exactly why Heckman corrections are needed, what they are, and how they substantively improve the story that could be told with a standard OLS technique. In essence, the authors assume that the readers are familiar with these kinds of models. Evidence for this conclusion appears in how they treat the comparison between the naïve model and the model that contains the selectivity correction. Nowhere do the authors state specifically how this model compares to the naïve model, in the context of sample selection bias. In other words, the authors do not state specifically and in non-technical jargon how this correction improves upon the naïve model. Moreover, I was not clear concerning why models 3 through 6 (Table 6) are needed and how they compare to models 1 and 2.

In the same vein, the authors conduct several sensitivity analyses (beginning on page 16), but do not explain exactly what these sensitivity analyses are. After reading the introductory paragraph that appears on page 16, I believe that the language is too technical for readers to understand why these analyses are needed and what they are specifically. The authors thus restrict the scope of consumers of article by their narrow and technical argot.

#### Conclusions, Caveats, and Substance

In the conclusion of the manuscript, the authors fail to tie the results back to the main impetus for the article: accessing the effects of measured versus unmeasured factors in selecting samples. While I understand the authors conclusion about weak effects of unobserved factors, they have done little to convince the reader of how these conclusions are informative for the research community at large. That is, if they see only modest differences in the parameter point estimates between the uncorrected and corrected models, then how does this important result help the research community in general? How does this result help analysts construct better samples? If these questions are not answered appropriately, then the exercises that the authors undertake diminish in substantive relevance.

#### Stylistic Comments

In tables, it may be helpful to use conceptual name of the concept being measured instead of the variable name as it appears on the statistical code. For example, they authors appear to be using the names of variables that appear in their code when they construct tables, especially table 4. I doubt this is conventional.

Table 4 is far too crowded. Overall, the formatting of the table needs to be changed.

Finally, after reading the article several times and examining the authors citations, I think that a more technical journal would be a better fit for this piece.