

Group decision modelling approach to analyse response behaviour of household travel survey: Examining proxy-response bias

Tomoki Kawano, Kenta Hosotani, and Takuya Maruyama

Social survey, such as travel survey, has shifted from interview-based to mail- and web-based surveys in contemporary studies. Owing to this change, the response rate to survey participation has been on the decline, and it is important to carefully examine the bias due to low response rate. While many types of biases exist, this study focus on the proxy-response bias in a household travel survey. Proxy response in household travel surveys can be defined as the situation where a household member responds to a travel survey for the entire household. This is in contrast with self-response, where each member of the target sample responds to their own survey. For example, a mother in a household may report based on guess the travel data of her entire household, which may include her father, husband, and children. In this case, the mother becomes a representative respondent for this household. In reality, her father may visit the hospital, take a walk to park, or go to community centre, and the mother may not be aware of these trips. Such crucial record is therefore inadvertently neglected, and thus the reported number of trips of the grandfather is underestimated. This is an example of proxy-response bias. These proxy response bias problems can be easily pointed out. However, as far as the authors know, no study has proposed the theoretical framework to analyse this problem. In this study, we propose a framework for the estimation of the probability that each member in household is a representative respondent. The group-based discrete choice framework provides an appropriate framework for this analysis.

To that end, the objectives of this study are to: 1) develop a group-based participation choice model of household travel survey, 2) examine the factors that influence response choice behaviour, and 3) demonstrate proxy response bias using developed model.

We assume that household i has J_i members and one-member j can be a representative respondent. The probability that household i responds through representative respondent j is defined as $P_{i,j}$. The probability that household i does not respond is $P_{i,0}$. If the utility function for a 'no response' is defined as $V_{i,0}$, and $V_{i,j}$ for response, the choice probability of each alternatives can be expressed by the following multinomial logit model.

$$P_{i,0} = \frac{\exp V_{i,0}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}}$$
$$P_{i,j} = \frac{\exp V_{i,j}}{\exp V_{i,0} + \sum_{j=1}^{J_i} \exp V_{i,j}} \quad \text{for } j \neq 0$$

This model is not the same as original multinomial logit model. This is because we can only observe the choice results for a part of the alternative; a representative respondent is generally unobservable. We can observe the data for the probability that households respond to survey $\sum_j P_{i,j}$, and we can estimate the

model with this partial observation. We maximize the likelihood L^* (or the log likelihood L) to estimate the parameters in the model as follows:

$$\begin{aligned}
L^* &= \prod_{i \in N} P_{i,0}^{1-\delta_i} \left(\sum_{j=1}^{J_i} P_{i,j} \right)^{\delta_i} \\
L &= \ln L^* \\
&= \sum_{i \in N} \left\{ (1 - \delta_i) \ln P_{i,0} + \delta_i \ln \sum_{j=1}^{J_i} P_{i,j} \right\} \\
\delta_i &= \begin{cases} 1 & : \text{household } i \text{ respond} \\ 0 & : \text{otherwise} \end{cases}
\end{aligned}$$

We can now examine the proxy response bias using this model. Let A be the event that individual j in household i is representative respondent and let B be the event that household i responds. The proxy response probability P_{proxy} is then given as follows, using Bayes' theorem.

$$\begin{aligned}
P(B) &= \sum_{j=1}^{J_i} P_{i,j}, P(A) = P_{i,j} \\
P(\bar{A}) &= 1 - P(A) = 1 - P_{i,j} \\
P(\bar{A} \cap B) &= P(B) - P(A) \\
&= \sum_{j=1}^{J_i} P_{i,j} - P_{i,j} \\
P(\bar{A}|B) &= \frac{P(\bar{A} \cap B)}{P(B)} \\
&= \frac{\sum_{j=1}^{J_i} P_{i,j} - P_{i,j}}{\sum_{j=1}^{J_i} P_{i,j}} \\
&= P_{\text{proxy}}
\end{aligned}$$

We used data from a household travel survey—referred to as Kumamoto Person Trip (PT) survey—conducted in 2012 in Kumamoto, Japan. We have basic household information of part of the target household ($n = 13,279$) including the age and gender of the non-responding household. These individual attributes and residential characteristics are explanatory variables in the proposed model. We have successfully estimated the statistical significant parameters with reasonable sign for the model.

We found that the participants were in a small-sized household (with respect to the number of members) located in an exclusive residential area. It was also observed that the representative participants were males over the age of 60 years and female from 40 to 60 years of age. Compared to individual-based choice model, ρ^2 is improved in a group-based choice model, which is one advantage of group-based

model over individual-based model.

To examine the proxy response bias, we divided the individual sample into two classes: self-response class and proxy-response class. The self-response class consists of individuals with 0% to 50% probability of proxy-response, whereas the proxy-response class consists of individuals with 50% to 100% probability of proxy-response. The difference between the two classes indicates a proxy-response bias. We found that the proxy-response class had less trip rates (average number of trips per day) for males of all ages, and large gap in the trip rates for females of the younger generation. Childcare and shopping trips would naturally be misreported by proxy-response and this may account for this gap. We also found a large gap in the trip rates for elderly non-working males, who may engage in walking and recreational trips that would be misreported by the proxy-response class.

In summary, this study developed a group-based choice model for response behaviour and a novel methodology for formulating the probability distribution of proxy response. We believe this is the first attempt at providing a theoretical framework and empirical analysis of proxy-response. This method will be valuable not only in travel surveys but also in the research and practice in the field of social survey in general.