Ministerie van Infrastructuur en Milieu

Estimating the nonresponse bias through Modeling of Nonresponse Behavior for the MPN

Raymond Hoogendoorn, Eline Scheepers, Olga Huibregtse & Sascha Hoogendoorn - Lanser

Netherlands Institute for Transport Policy Analysis

Delft, 14 September 2016

Contents

- Introduction;
- Method;
- Results;
- Discussion;

Introduction

- Mobility patterns are influenced by many factors (e.g., life events, changes in the use of ICT, etc.);
- Insight into this influence is typically derived from longitudinal panel data;
- Netherlands Mobility Panel;
- Accuracy is influenced by nonresponse bias;
- Initial nonresponse, question or trip diary nonresponse and attrition;

Introduction[2]

- Nonresponse may, but need not induce a nonresponse bias;
- For example: Suppose that between two waves the government implemented measures to stimulate working from home;
- From the location-based diary one could conclude that the number of trips has decreased;
- However, analyses of attrition show that a substantial number of respondents with a higher education have dropped out;
- Correlation between education level and the number of trips;
- Bias

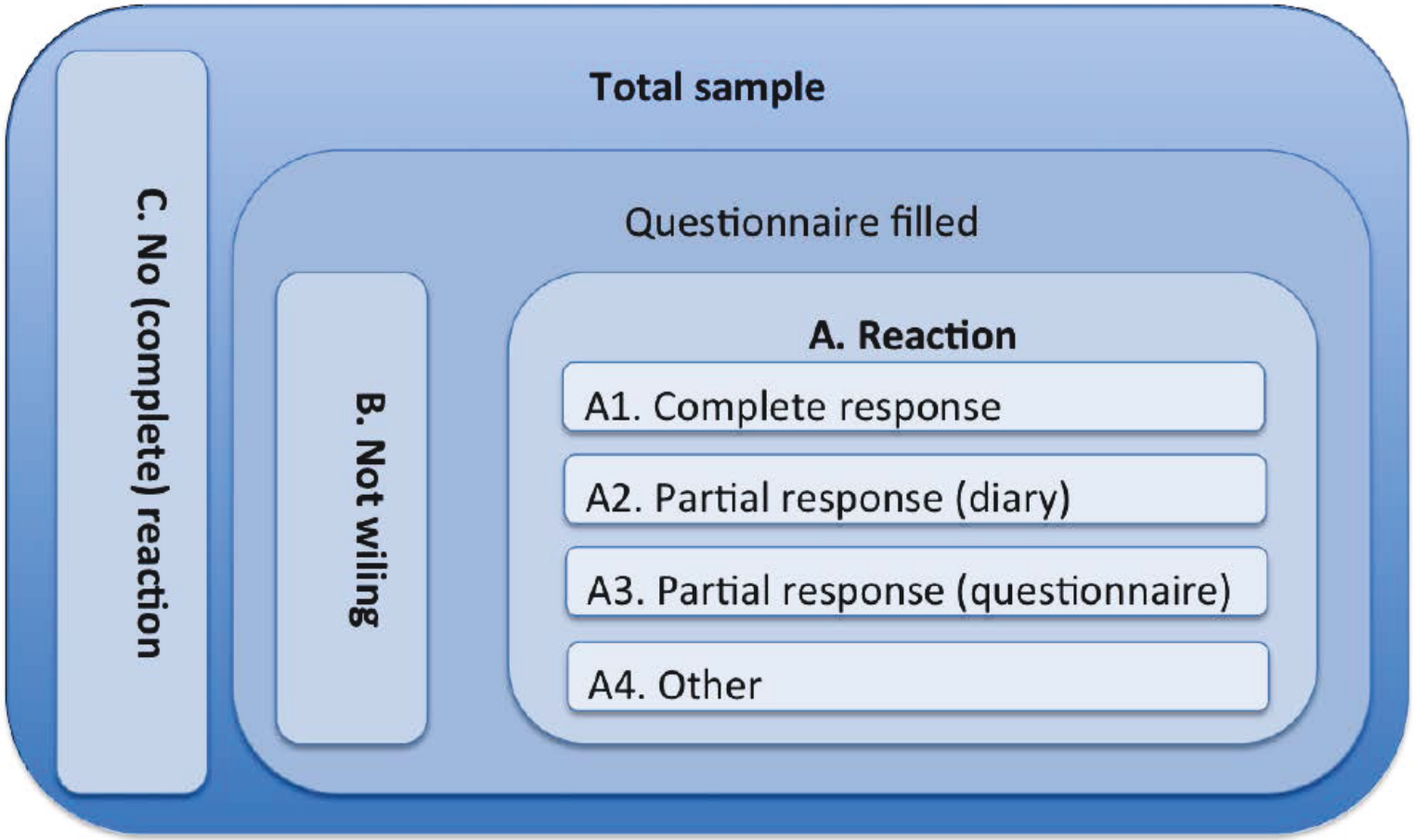
Introduction[3]

- Insight is needed into the willingness to participate;
- Research objective: to gain insight into nonresponse bias through modeling of the willingness to participate in a household panel;
- Allows for empirically underpinned insight into and quantification of the different factors that influence nonresponse behavior and the nonresponse bias;
- Insight into how to best model the willingness to participate in a household panel;
- More accurate knowledge on mobility choices and the relationship with various factors, such as life events and developments in ICT use;

Method

- Research questions:
  - What is the magnitude of the nonresponse and willingness to participate in the household panel;
  - Is there a significant difference in personal and household characteristics between the individuals who indicated to be willing to participate versus individuals who indicated that they were not willing to participate or do not react at all?
  - How can the willingness to participate in a household panel best be modeled?
  - Is a nonresponse bias present and what is the magnitude of this bias?

Method[2]


Total sample
C. No (complete) reaction
Questionnaire filled
B. Not wiling
A. Reaction
A1. Complete response
A2. Partial response (diary)
A3. Partial response (questionnaire)
A4. Other

Method[3]

- Chi-square tests with a confidence level of 95%;
- Ordinal logit model using Biogeme Python;
- Nonresponse bias: $\text{bias}(\hat{\bar{y}}_{un}) \approx \bar{\phi}^{-1}\sigma_{\phi}\sigma_{y}\rho_{\phi,y}$
- "How often do you use your car (either as a driver or as a passenger)?"
- Spearmans's rho;

Method[4]

- Ordinal logit model:

$$\begin{aligned}\text{logit}(p_1 + p_2 + p_3) &= \log \frac{p_1 + p_2 + p_3}{1 - p_1 - p_2 - p_3} \\ &= \alpha + \beta_1 X_{\text{gen}} + \beta_2 X_{\text{age}} + \beta_3 X_{\text{edu}} + \beta_4 X_{\text{emp}} + \beta_5 X_{\text{Hsit}} + \beta_6 X_{\text{Hsiz}} \\ &+ \beta_7 (X_{\text{gen}} X_{\text{age}}) + \beta_8 (X_{\text{gen}} X_{\text{hsit}}) + \beta_8 X_{\text{wav}}\end{aligned}$$

- Main factors and interactions with gender;

Results

- Magnitude of nonresponse:
  - 44.2% were willing to participate in the study (A);
  - 26.9% were not willing to participate in the studt (B);
  - 28.8% did not (fully) respond;

- Relationship with personal and household characteristics:

Characteristic	Value	*df*	*p*-value
Gender	**387.122**	**2**	**.000**
Age	**437.014**	**12**	**.000**
Education	**89.135**	**4**	**.000**
Employment status	**338.380**	**10**	**.000**
Household situation	**251.528**	**6**	**.000**
Household size	**165.999**	**10**	**.000**
Wave	**118.087**	**2**	**.000**

Results[2]

- Parameter values ordinal logit model;
- Interpretation is non straightforward;
- Sign and relative magnitude;
- Several significant variables;
- Examples;

Parameter	Value	Std. error	CI 2.5%	CI 97.5%
Gender				
Male	**-.438**	**.080**	**-.596**	**-.280**
Age				
18-24	-.127	.085	-.295	.039
25-34	**.356**	**.090**	**.178**	**.533**
35-44	**.507**	**.104**	**.302**	**.713**
45-54	**.475**	**.110**	**.259**	**.692**
55-64	**-.075**	**.112**	**-.146**	**-.001**
Education				
No - low	**.191**	**.044**	**.105**	**.278**
Medium	**.194**	**.040**	**.116**	**.272**
Employment status				
Paid	**-.257**	**.070**	**-.394**	**-.119**
Disabled	**-.352**	**.108**	**-.565**	**-.140**
Retired	-.093	.075	-.242	.055
Student	**-.317**	**.073**	**-.461**	**-.173**
Unemployed	**-.350**	**.103**	**-.553**	**-.147**
Household situation				
Single	-1.659	1.176	-4.687	.448
With young child 0-13	**.326**	**.075**	**-.474**	**-.178**
With young child 13-17	-.089	.077	-.241	.061
Household size				
Size 1	1.628	1.179	-.487	4.660
Size 2	**-.368**	**.112**	**-.588**	**-.148**
Size 3	-.145	.105	-.352	.062
Size 4	.006	.102	-.193	.207
Size 5	.023	.110	-.240	.193
Wave				
Wave 2	-.006	.051	-.108	.093
Thresholds				
τ_1	-.851	.134	NA	NA
τ_2	.336	.134	NA	NA

Results[3]

- Interactions
- Several significant interactions with gender;
- Examples;

Parameter	Value	Std. error	CI 2.5%	CI 97.5%
Gender: Age				
Male: Age 18-24	**.554**	**.095**	**.367**	**.745**
Male: Age 25-34	**-.333**	**.100**	**-.531**	**-.136**
Male: Age 35-44	**-.781**	**.116**	**-1.010**	**-.553**
Male: Age 45-54	**-.840**	**.115**	**-1.066**	**-.614**
Male: Age 55-64	**-.419**	**.123**	**-.660**	**-.177**
Gender: Household situation				
Female: Single	**-.808**	**.084**	**-.973**	**-.643**
Female: With young child 0-13	**.335**	**.098**	**.143**	**.527**
Female: With young child 13-17	-.011	.108	-.223	.200

Results[4]

- Estimated probabilities;
- Very close to the observed percentages;
- Observed willingness was 44.2% versus 26.9 and 28.8% for not willing and no reaction;

Willingness	N	Mean	Std	Min	Max
Reaction	14679	.438	.070	.248	.602
Not willing	14769	.269	.011	.225	.281
No reaction	14679	.291	.061	.171	.486

Results[5]

- Nonrespons bias:
- Recall: $\text{bias}(\hat{\bar{y}}_{un}) \approx \bar{\phi}^{-1}\sigma_\phi\sigma_y\rho_{\phi,y}$
- Spearman's rho for car uses amounts to .21;
- This means that higher probability of willingness is accompanied by less frequent car use;
- Identification of bias;
- Bias amounts to .033;

Discussion

- Approximation of nonresponse bias through modeling of willingness to participate;
- Close resemblance of estimated probabilties with observed willingness to participate;
- Ordinal logit model performs quite well;
- Various personal and household characteristics are significantly related to willingness;
- Identification of bias;

Discussion[2]

- However, only first two waves were included;
- Is a nested logit model better?
- How to develop a correction factor for location based trip diaries?

Questions?