The 23rd summer course:
Behavior Modeling in Transportation Networks
September 11-13, 2024

Bayesian Truth Serum and Stated Preference Survey

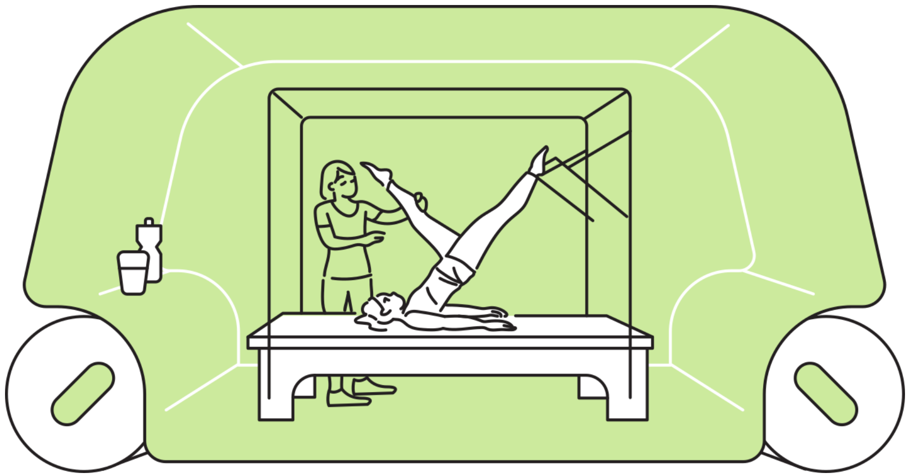
Yusuke Hara (Tohoku University)

TOHOKU UNIVERSITY

This is a joint work with Tomoki Nishi in TOYOTA CRDL, INC.,

Background

- **Demand forecasting for unknown goods and services**
  - If autonomous driving becomes a reality, will you buy an autonomous vehicle?
  - If you buy an autonomous vehicle, will you live in the city center or in the suburbs?
  - Do you want to use an autonomous mobile gym that comes to your home?



Stated Preference (SP) Survey

- Preference data that observe preference in a hypothetical situation is called **Stated Preference (SP)**.
    - Specifically, discrete choice data is referred to as stated choice (SC).

- SP surveys enable us to forecast the demand for new transportation services that do not currently exist.

- Differences from questionnaire surveys
    - Controls for the effects of trade-offs between attributes of alternatives based on experimental design.
    - responses are used to estimate behavioral models.

Example of Stated Preference Survey

- Which transport mode will you use?

	Subway	Bus	LRT
Total travel time	25 min	40 min	30 min
Fee	220 JPY	200 JPY	250 JPY
Access time	8 min	2 min	5 min
Egress time	5 min	1 min	4 min
Frequency	10 per hour	6 per hour	5 per hour
Choice	◉	○	○

Example of Stated Preference Survey

- Which transport mode will you use?

	Subway	Bus	LRT
Total travel time	38 min	22 min	17 min
Fee	400 JPY	250 JPY	350 JPY
Access time	2 min	5 min	4 min
Egress time	3 min	7 min	8 min
Frequency	5 per hour	8 per hour	4 per hour

Choice ○ ● ○

By controlling for the trade-offs between the attributes of each option,
SP survey enable us to estimate sensitivity with respect to each attribute.

Hypothetical bias in SP survey

1. Experimental Scenario Uncertainty

- Ambiguity and uncertainty about unfamiliar goods and services
- Validity of hypothetical scenarios

2. Heterogeneity of respondents

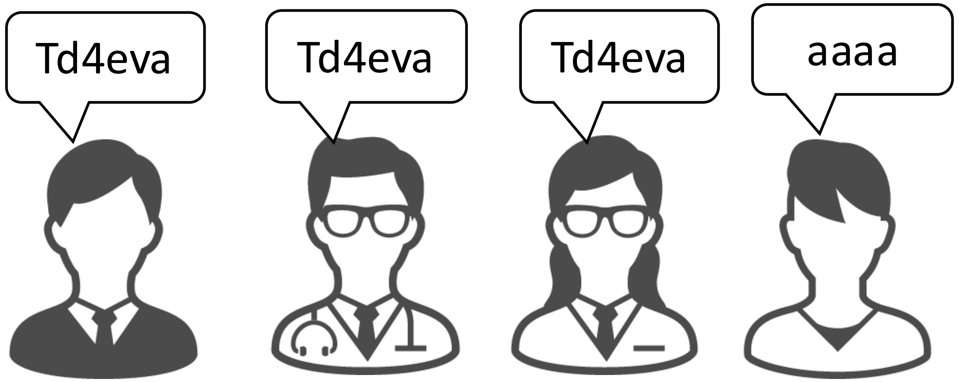
- Information and knowledge possessed by each respondent
- Preference heterogeneity

3. Dishonest response

- Survey credibility, including policy maneuvering bias and justification bias
- Survey stability, including response fatigue and cold start issues

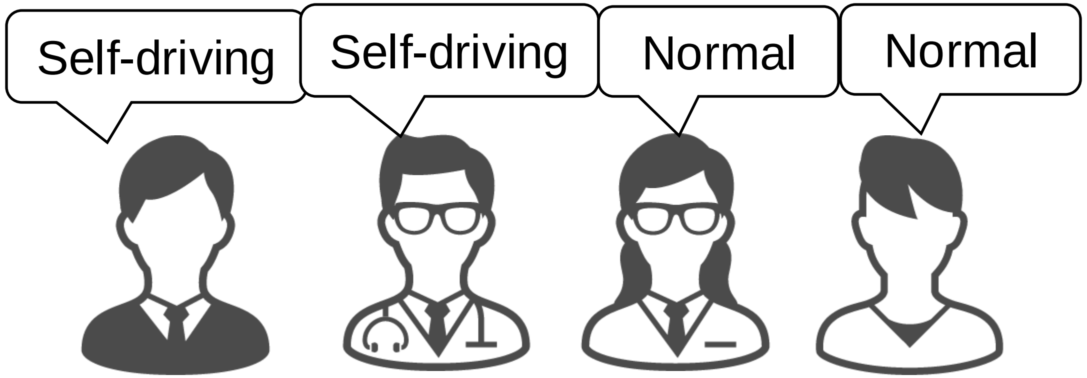
Method to eliminate inaccurate responses

- In contrast to the quality control of **crowdsourcing** in the field of human computation



Crowdsourcing

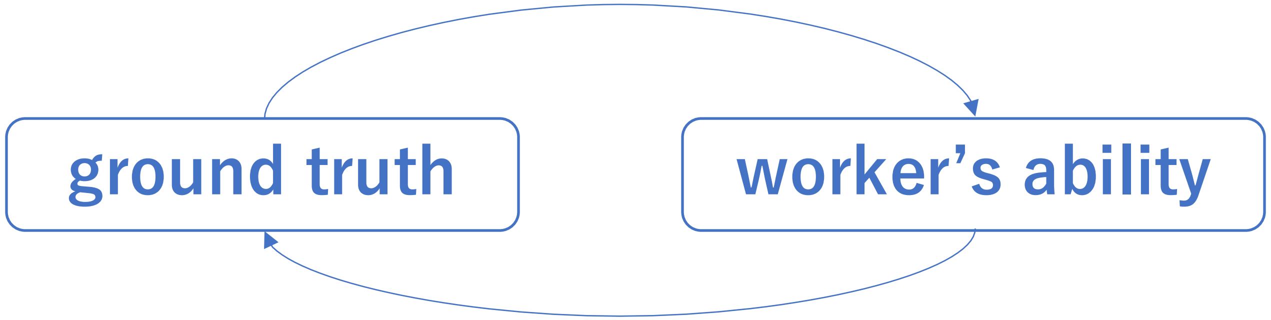
Do you want a self-driving car or a normal car?



Stated Preference

Spam worker detection in crowdsourcing

- Detect spam workers by calculating the percentage of correct answers.
- Detect spam workers by having them solve the same problem and deciding the answer by majority vote
  - Both are inefficient.
- Latent class model (Dawid and Skene, 1979; Boxall and Adamowicz, 2002) can estimate the workers' ability and ground truth simultaneously.

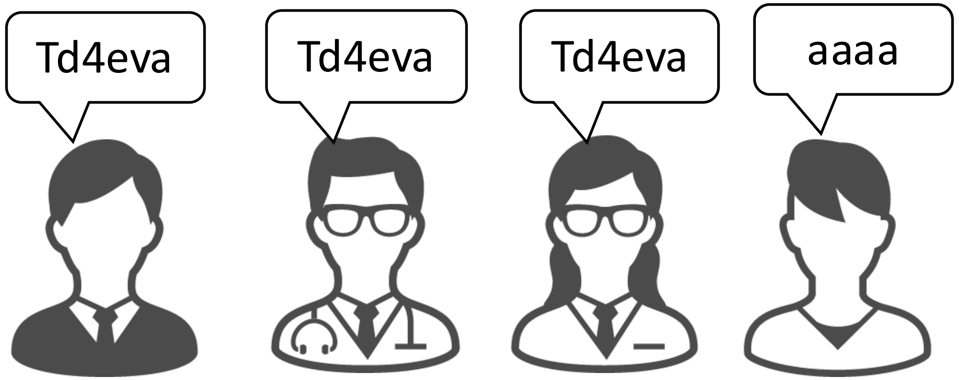


Classification of workers' skills can lead to detection of spam workers.

EM algorithm

However, there is no ground truth in the case of SP

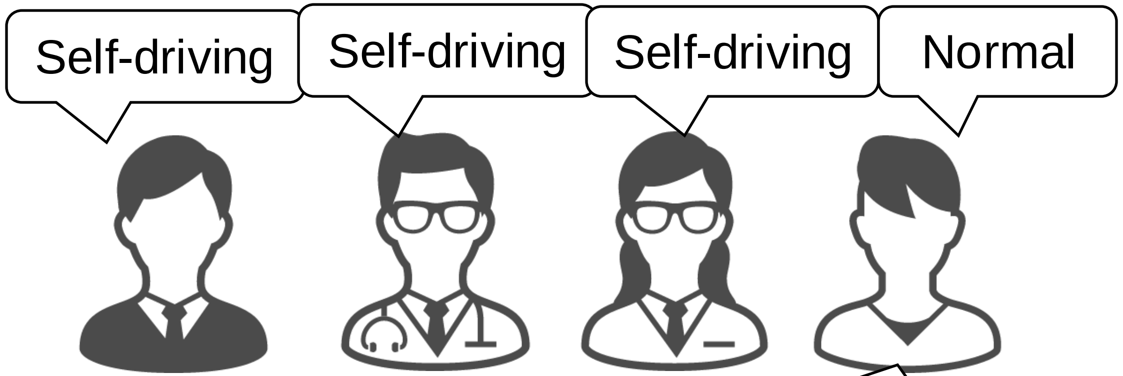
Crowdsourcing



Detected as spam

Stated Preference

Do you want a self-driving car or a normal car?



We cannot determine whether a response is a spam response (a random response) or a personal preference, even if it is in the minority.

The method to induce information and knowledge possessed by respondents

- Can we detect heterogeneity in respondent preferences and dishonest responses?

- **Bayesian Truth Serum** (Prelec, 2004)
  - A kind of proper scoring rule. (Johnson et al, 1990)
  - In this mechanism, responses are scored such that the highest score is obtained when the true subjective probability is answered.
  - BTS can be used to
    - Improve the accuracy of survey results
    - Identification of superior respondents
    - Behavioral change of respondents (truth-telling)
  - **We will apply the BTS to SP survey.**

Bayesian truth serum

- The original BTS is the mechanism design to make a survey to answer things that are difficult to answer under normal conditions.
  - e.g., Have you ever shoplifted?, Are you racist?

Q1: Any category question
Q2: Questions that make you predict "how others will respond to Q1."

- For example,
  - Q1: Have you ever shoplifted? Yes/No
  - Q2: How many people do you think would answer Yes to Q1? ____%

BTS Score

- Respondent i's response of Q1 is denoted by x_{ik} and the response of Q2 is denoted by y_{ik}.
- The BTS score is defined as

$$\overline{x_k} = \frac{1}{n}\sum_{i=1}^{n} x_{ik}$$

Arithmetic mean of actual responses

$$\log \overline{y_k} = \frac{1}{n}\sum_{i=1}^{n} \log y_{ik}$$

Geometric mean of predicted responses

$$BTS\ Score_i = \underbrace{\sum_k x_{ik} \log \frac{\overline{x_k}}{\overline{y_k}}}_{\text{Information score}} + \alpha \underbrace{\sum_k \overline{x_k} \log \frac{y_{ik}}{\overline{x_k}}}_{\text{Prediction score}}$$

BTS Score

Q1 Have you ever shoplifted?	Q2 How many people do you think would answer Yes to Q1?	BTS Score
Yes	20%	+0.31
No	10%	-0.18
Yes	5%	+0.09
No	30%	-0.09
・・・	・・・	・・・
No	25%	+0.32

"Good" response

"Poor" response

Percentage of Yes
25%

Predicted percentage of Yes
18%

BTS Score

$$BTS\ Score_i = \underbrace{\sum_k x_{ik} \log \frac{\overline{x_k}}{\overline{y_k}}}_{\text{Information score}} + \alpha \underbrace{\sum_k \overline{x_k} \log \frac{y_{ik}}{\overline{x_k}}}_{\text{Prediction score}}$$

- Information score is
  - If the "Actual Percentage" is higher than the "Predicted Average", those who chose the option will receive a high score.
  - This is the rule that the majority gets a higher score compared to everyone else's prediction.
- Prediction score is
  - The closer the prediction is to the actual percentage, the higher the score.

The characteristics of BTS

- **Scoring does not require an external ”ground truth”**
  - No need for verification that the person is a shoplifter.

- **Scoring independent of response distribution**
  - Possibility of high scores even for minority opinions
  - It is often the case that a small group of people with some expertise, or a group of actual criminals, know more about the real situation than the general public.

- **Incentive compatibility**
  - Linking BTS scores to incentives can elicit desired behavior (truth-telling)
  - To increase the BTS score, it is incentive compatible to answer honest choices and true subjective probabilities.

Research idea: BTS-SP + Latent class

- Detect dishonest spam respondents using both choice and predictive responses, and continuously separate spam respondents from those useful for model estimation.

- BTS scores accurately identify spam responses and responses due to preference heterogeneity.

- In doing so, we improve the predictive performance of the model and clarify the responses to the important variables.

The difference from experimental design of SP survey

- Which transport mode will you use?

	Subway	Bus	LRT
Total travel time	25 min	40 min	30 min
Fee	220 JPY	200 JPY	250 JPY
Access time	8 min	2 min	5 min
Egress time	5 min	1 min	4 min
Frequency	10 per hour	6 per hour	5 per hour

Choice ◉ ○ ○

- What percentage of people do you think would make the same choice you did?

Prediction | 80 % |

Just add this!

Approach of our analysis

Observed data
from BTS-SP survey

responses (x_{ik}, y_{ik}), choice and prediction

Make a choice model: $P(x_{ik})$ from x_{ik}

Make a prediction model: $P(y_{ik})$ from y_{ik}

Calculate a pseudo BTS score (pseudo information score and pseudo prediction score) for each respondent: IS_i, PS_i

Two stage latent class model using pseudo BTS scores

membership function using IS_i, PS_i

Subject group

No-spam

Spam

membership function using
socioeconomic characteristics

choice model of class 1

choice model of class 2

random response model

Pseudo BTS score

- BTS score

$$\overline{x_k} = \frac{1}{n}\sum_{i=1}^{n} x_{ik}$$

Arithmetic mean of actual responses

$$\log \overline{y_k} = \frac{1}{n}\sum_{i=1}^{n} \log y_{ik}$$

Geometric mean of predicted responses

$$BTS\ Score_i = \sum_k x_{ik} \log \frac{\overline{x_k}}{\overline{y_k}} + \alpha \sum_k \overline{x_k} \log \frac{y_{ik}}{\overline{x_k}}$$

- Pseudo BTS score

$$\overline{x_k} = \frac{1}{n}\sum_{i=1}^{n} \hat{P}(x_{ik})$$

Arithmetic mean of actual responses by using a choice model

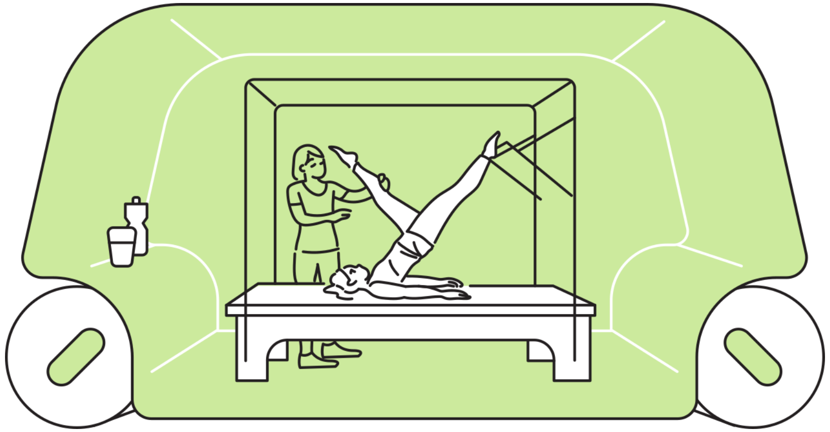
$$\log \overline{y_k} = \frac{1}{n}\sum_{i=1}^{n} \log \hat{P}(y_{ik})$$

Geometric mean of predicted responses by using a prediction model

$$pseudo\_BTS\ Score_i = \sum_k x_{ik} \log \frac{\overline{x_k}}{\overline{y_k}} + \alpha \sum_k \overline{x_k} \log \frac{y_{ik}}{\overline{x_k}}$$

An example of BTS-SP survey

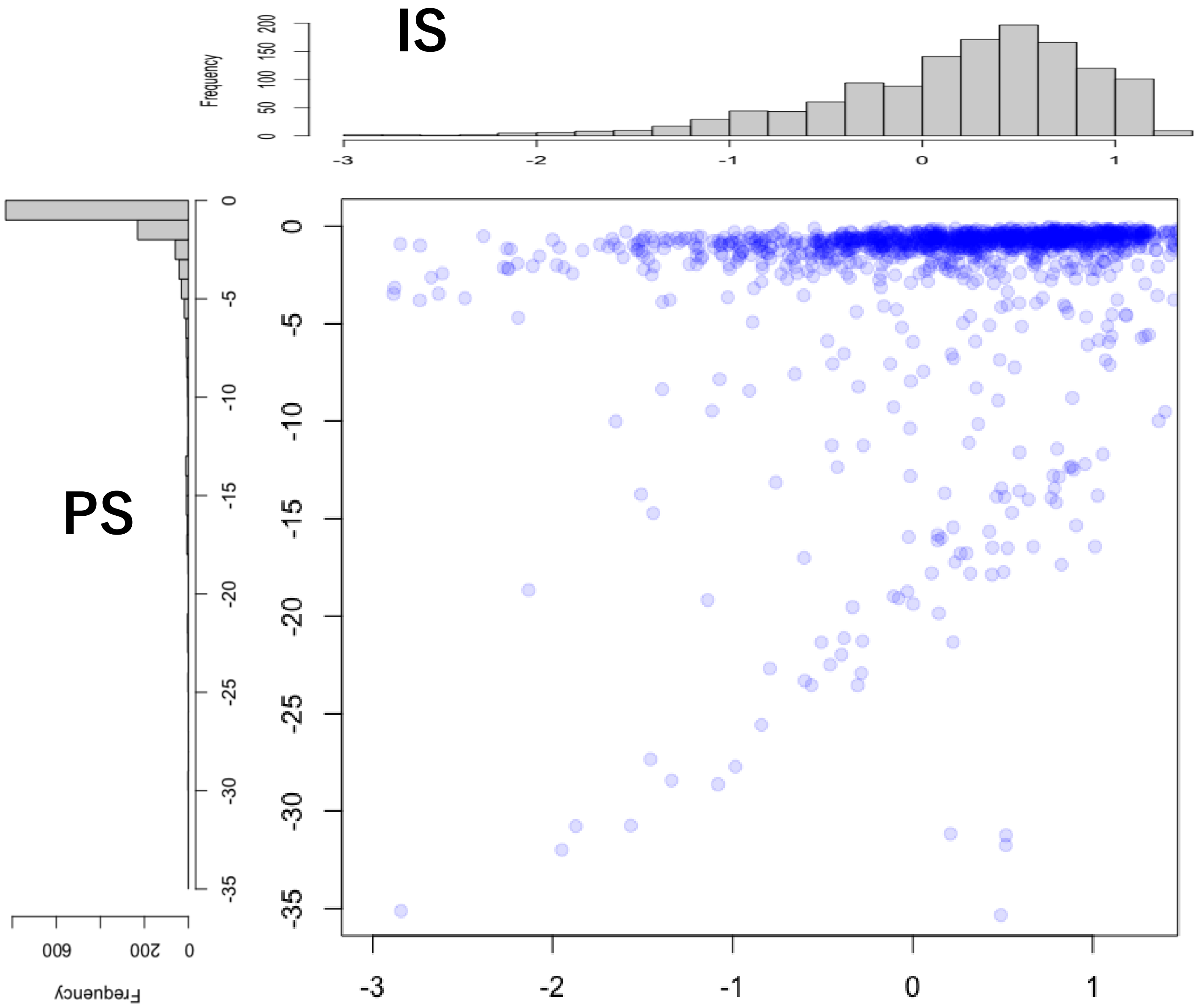
- Demand forecasting for mobile gym



VS online gym, gym

- Attributes of each option
  - Availability of personal trainer
  - Monthly Fee
  - Distance from home
  - Business Hours
  - Availability of swimming pools
  - Availability of parking

The result of pseudo BTS score

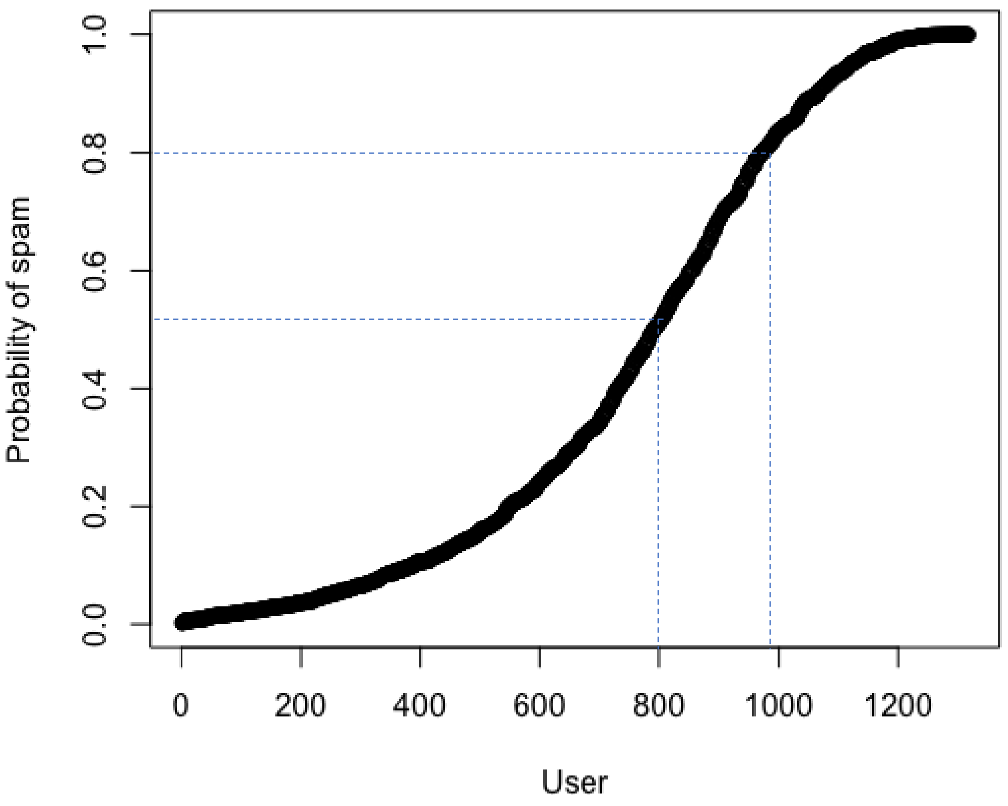


- High IS = Users for whom models are "easy to guess".
- Low IS = Users for whom the model is "hard to guess".

- PS variation is significantly greater than IS variation.
- Low IS does not necessarily mean low PS.
- They are not easily correlated.

Detection of spam respondents

Estimation of spam respondent probability for each user



Membership function of spam respondents

$$P_{i,spam} = \frac{\exp(\beta_1 \cdot x_{i,IS} + \beta_2 \cdot x_{i,PS} + \beta_3)}{1 + \exp(\beta_1 \cdot x_{i,IS} + \beta_2 \cdot x_{i,PS} + \beta_3)}$$

Parameter	estimates
IS	4.804
PS	0.0044
constant	-0.824

- Spam respondent probability tends to increase with lower PS and decrease with higher PS.

Improvement of model performance

- Model performance

	MNL	Latent MNL	Pseudo BTS Latent MNL
# of parameters	15	27	31
# of observations	7902 (1317)	7902 (1317)	7902 (1317)
Initial LL	-5477.249	-5477.249	-5477.249
final LL	-4456.009	-4384.219	-3935.282
likelihood ratio	0.184	0.195	0.276

- Number of model parameters is almost the same as the normal latent class model (+4), but model performance is greatly improved.
- The difference is created by the pseudo BTS score, which scores the responses of each subject in the population.
  - Sort "honest respondents" who respond to attributes from "spam respondents" who do not.

Summary

- Demand forecasting for unknown goods and services remains an important challenge.

- We proposed a new experimental design, the BTS-SP survey, to overcome the problems of classical SP surveys and to detect dishonest responses.

- Two-level latent class model estimation using pseudo BTS score.
  - Significantly improved model performance over naïve latent class models by identifying the preference heterogeneity and detecting spam respondents.

References

1. Dawid, A. P., & Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(1), 20-28.
2. Boxall, P. C., & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and resource economics*, *23*(4), 421-446.
3. Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science*, *306*(5695), 462-466.
4. Johnson, S., Pratt, J. W., & Zeckhauser, R. J. (1990). Efficiency despite mutually payoff-relevant private information: The finite case. *Econometrica: Journal of the Econometric Society*, 873-900.
5. Menapace, L., & Raffaelli, R. (2020). Unraveling hypothetical bias in discrete choice experiments. *Journal of Economic Behavior & Organization*, *176*, 416-430.
6. Beck, M. J., Fifer, S., & Rose, J. M. (2016). Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. *Transportation Research Part B: Methodological*, *89*, 149-167.
7. Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it?. *Transportation research part A: policy and practice*, *61*, 164-177.
8. Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H., & Lancsar, E. (2021). Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. *Journal of choice modelling*, *41*, 100309.
9. Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H., & Lancsar, E. (2021). Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. *Journal of choice modelling*, *41*, 100322.
10. Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: methodological*, *44*(6), 735-752.