Distances in Latent Space: A Novel Approach to Analyzing Conjoints

Simon Hoellerbauer

University *of* North Carolina *at* Chapel Hill

Research Objectives

Methodological objective:

- Merge IRT model and logistic regression to estimate the effect of distances in latent space on outcomes of interest in conjoint survey experiments

Substantive problem:

- Do individuals feel closer to more local organizations and do feelings of closeness make individuals more likely to want to interact with an organization?

Motivation

- Recent work on conjoints has underlined the need to reconsider how we analyze them [2], with one proposal to use an IRT framework [1].
- I adapt this framework to focus on the actual distance between profiles and respondent and its effect.

Two Part Model

1) IRT component to estimate positions in latent space:

$$\Pr(Y_{ik} = 1|\mathbf{x}_{ik1}, \mathbf{x}_{ik2}) = \Phi(b(\mathbf{x}_{ik1}, \mathbf{x}_{ik2})\theta_i - g(\mathbf{x}_{ik1}, \mathbf{x}_{ik2}))$$

where $b(\mathbf{x}_{ik1}, \mathbf{x}_{ik2}) = 2(x_{ik1} - x_{ik2})^\top \boldsymbol{\beta}/\sigma$ and $g(\mathbf{x}_{ik1}, \mathbf{x}_{ik2}) = \boldsymbol{\beta}^\top(\mathbf{x}_{ik1}\mathbf{x}_{ik1}^\top - \mathbf{x}_{ik2}\mathbf{x}_{ik2}^\top)\boldsymbol{\beta}/\sigma$.

2) Logistic regression to estimate effect of distances:

$$\Pr(W_{ij} = 1|\mathbf{x}_{ij1}, \mathbf{x}_{ij2}) = \text{logit}^{-1}(\gamma_0+\gamma_1 * (2\theta_i(\mathbf{x}_{ij1} - \mathbf{x}_{ij2})^\top\boldsymbol{\beta}+ \boldsymbol{\beta}^\top(\mathbf{x}_{ij2}\mathbf{x}_{ij2}^\top - \mathbf{x}_{ij1}\mathbf{x}_{ij1}^\top)\boldsymbol{\beta})$$

where Y_{ik}, W_{ij} are 1 if profile 1 in pairs k and j (different pairs used for each outcome) for respondent i is chosen, and 0 if not, and $\mathbf{x}_{ik1}, \mathbf{x}_{ik2}$ and $\mathbf{x}_{ij1}, \mathbf{x}_{ij2}$ represent the attributes of profiles 1 and 2 in pairs k and j, respectively, for respondent i.

Note that the term with the γ_1 coefficient is equal to $\theta_i - \mathbf{x}_{ij2}^\top\boldsymbol{\beta})^2 - (\theta_i - \mathbf{x}_{ij1}^\top\boldsymbol{\beta})^2$: the *difference in the distance between ideal points and profile locations*: positive = i closer to profile 1 than profile 2.

Empirical Application: Localness and Organizations

- 676 students at UNC - Chapel Hill completed conjoint survey experiment
- Attributes made organizations more or less student-like (demographically local) or geographically local
- Each saw 15 randomly created pairs of hypothetical vote registration organizations
- Students were asked two outcome questions:
  1. W: Would you be more likely to attend a meeting held by organization 1 or organization 2?
  2. Y: With which organization would you say you feel more of a personal connection?
- Used profiles 2-15 for IRT portion; profile 1 for logistic regression component
- Models were estimated using `Stan`'s `R` interface `rstan`

Results

Fig 1: $\hat{\beta}$—Determine Org. Locations (Posterior medians with 95% cred. int.)

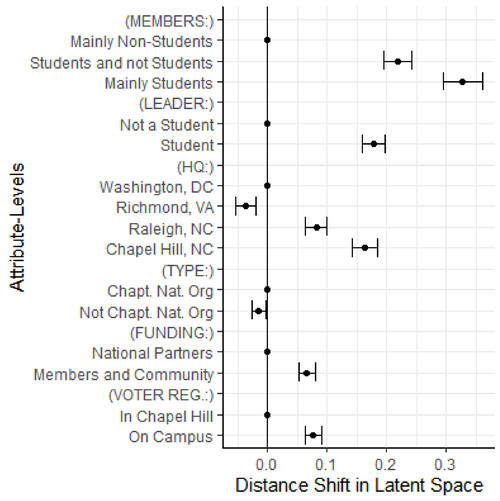


Fig 2: $\hat{\theta}_i$'s And All Possible $\mathbf{X}\hat{\beta}$ (Posterior medians with 95% cred. int.)

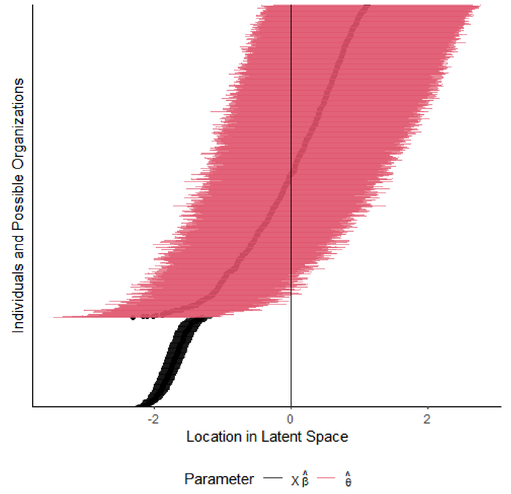
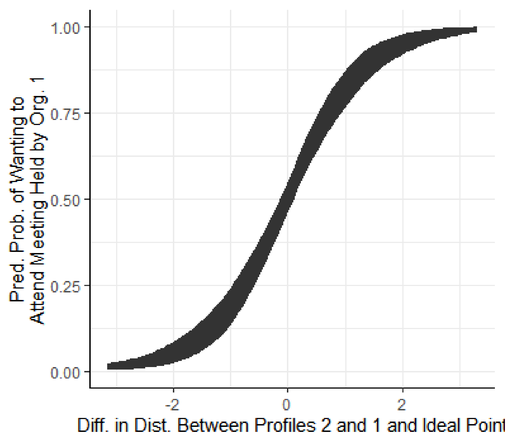


Fig 3: Effect of Diff. in Dist. Between Ideal Points and Profiles on Probability of Wanting to Attend Meeting



- Local organizational traits moved organizations closer to positive pole of latent space (Fig. 1)
- 59.3% of students had estimated positions with credible intervals entirely to the right of the most student/local organization (Fig. 2)
- As difference in distances increases—respondent is closer to org. 1 than org. 2—the probability of wanting to attend a meeting held by org. 1 increases (for reference, within data distances were normally distributed around 0, with standard deviation 1) (Fig. 3)

Assessing Model Fit

- Use Area Under the ROC Curve (because Bernoulli distributed outcomes)
- Use profiles 2-15 to assess fit of logistic regression portion and profile 1 to assess fit of IRT portion

AUCs (Posterior Medians with 95% Credible Intervals)

IRT:	0.868 [0.858, 0.878]
Logistic Regression:	0.765 [0.762, 0.769]

- Logistic regression part of model does not fit as well; it is possible that students took other factors into account besides distance, or that the form of the distance is different (absolute difference, for example)

Conclusion

- **Substantively**, students were more likely to want to attend a meeting of the organization to which they were closer in the latent space—this shows that closeness matters for engagement
- **Methodologically**, this project demonstrates the value of thinking of conjoint profiles and individuals as being located in a latent space

Next Steps

- Restructure experiment so that second question (W) is not forced-choice but asked about each profile in turn; this can get a better estimate of the effect of distance; requires modification of logistic regression portion of model
- Simulation study to investigate approach more fully

Email: hoellers@unc.edu

Full presentation slides available **here**.

References

[1] Devin Caughey, Hiroto Katsumata, and Teppei Yamamoto. Item response theory for conjoint survey experiments. Working Paper, 2019.

[2] Scott Abramson, Korhan Koçak, and Asya Magazinnik. What do we learn about voter preferences from conjoint experiments? Working Paper, 2019.