INSTITUTE FOR SOCIAL RESEARCH
PROGRAM IN SURVEY AND DATA SCIENCE
UNIVERSITY OF MICHIGAN

Improving Inferences Based on Survey Data Collected Using Mixed-Mode Designs

Wenshan Yu, Michael Elliott, Trivellore Raghunathan

May. 13th

Motivation

Mixed-mode designs have become more common for survey data collection

Advantages

- Improve sample balance
- Save survey costs

Disadvantage

- Introduce mode-specific measurement error

How best to analyze the pooled data collected with mixed-mode designs?

References: De Leeuw (2018)

Context: Arab Barometer Study

- Largest repository of public opinion data in the Middle East and North Africa region
  - 15 countries: Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Palestine, Saudi Arabia, Sudan, Tunisia, and Yemen
  - Six waves since 2006
  - Led by Princeton University and supported by several other research institutes
- Switched from FTF to mixed modes (FTF and TEL) in 2020 (Jordan, Tunisia, and Morocco)
- Measuring regime support in authoritarian countries can be sensitive
- Mixed evidence on mode effects due to social desirability bias between TEL and FTF
- In this case, how to make inferences to incorporate the potential mode effects?

References: Robinson and Tannenberg (2019); Holbrook, Green and Krosnick (2003); Klausch, Hox, and Schouten (2013)

Literature review

Approach #1: Calibrate mixed-mode data to approximate what would have been collected with a single mode

- Impute at the individual level for the unpreferred mode
- E.g., Powers, Michra, and Young (2005); Elliott et al. (2009); Kolenikov and Kennedy (2014); Park, Kim, and Park (2016)
- Assume mode effects are present and we know the preferred mode

Approach #2: Develop mode-specific estimate level weights

- Optimize the combined estimate in terms of MSE (Suzer-Gurtekin et al., 2012 and 2013)
- Fix weights of mode-specific estimates in longitudinal surveys to stabilize mode-related measurement error (Buelens and Van den Brakel, 2015)
- Do not aim to tease out mode effects

We propose new approaches to incorporate potential mode effects!

Setup

- A normally distributed outcome Y; interested in inferring its population mean (θ).
- Borrow from the causal inference literature: only observe marginal means (u_a and u_b) and marginal variances (σ_a^2 and σ_b^2) corresponding to the observed mode
  - The outcome under the unobserved mode is considered as a “counterfactual”
  - Never jointly observe y_{ai} and y_{bi}, thus impossible to estimate the correlation
  - Assume the correlation as 0 for the proposed approaches for simplicity
- Assume at least one mode provides unbiased estimate of the population mean
- Consider two settings: 1) when having prior info on the preferred direction and 2) when there is no info on the preferred direction
  - “Preferred directions” reflect people’s beliefs on which direction is less biased
  - I.e., whether a smaller or a larger estimate is less biased
- Estimation strategy: $\theta = \min(u_a, u_b)$ or $\max(u_a, u_b)$ or $\frac{u_a+u_b}{2}$

References: Pattanayak et al. 2012; Li et al. 2021.

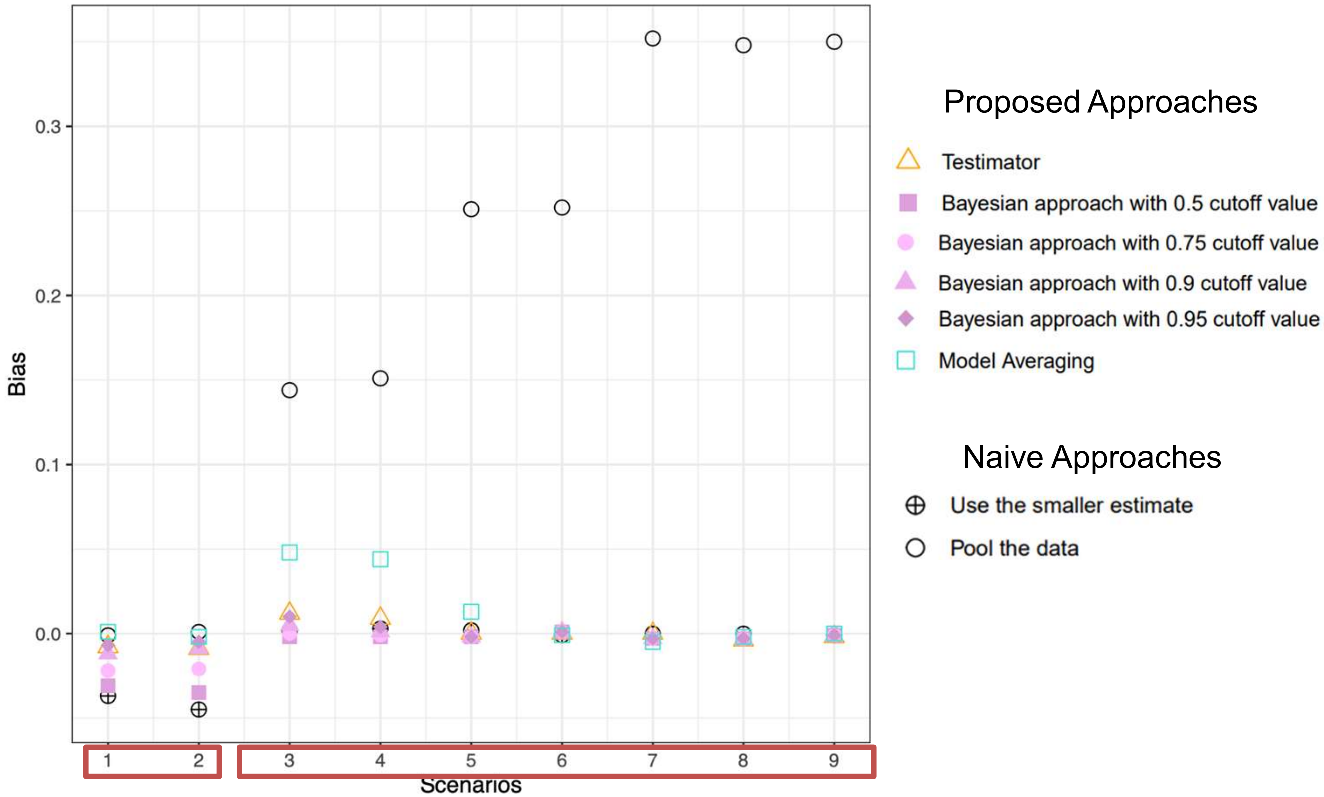
Three approaches (prior info available)

1. Testimator
    - Frequentist model selection
    - First test the equality of variances, then test the equality of means
    - Pool estimates of means/variances if fail to reject the null hypothesis
    - If reject the null hypothesis, take the estimate in the preferred direction
2. Bayesian
    - Bayesian model selection
    - First use estimates of an effect size to determine whether the means are the same
    - Then draw population mean from pooled data if evidence suggests equal means
        - Consider cutoff values to be 0.5, 0.75, 0.9, and 0.95
    - If evidence suggests different means, take the estimate in the preferred direction
3. Model averaging
    - Consider four models that differ on whether specifying different means and variances
    - Combine the four models with weights equal to the posterior probability that the given model is correct

Simulation Study

- Data generation model: $$\begin{pmatrix} y_{ai} \\ y_{bi} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \mu_b \end{pmatrix}, \begin{pmatrix} 1 & 0.95\sigma_b \\ 0.95\sigma_b & \sigma_b^2 \end{pmatrix} \right)$$
- Vary u_b and σ_b^2 to implement different level of mode effects.
- Assume mode A provides unbiased responses and mode B produces positive mode effects
- Have prior info that a smaller estimate is preferred
- Compare the three proposed approaches with two naïve approaches
  1) taking the smaller estimate of $\bar{y}_a$ and $\bar{y}_b$
  2) pooling the data
- Sample size: 500 (250 per mode), # of simulation: 500

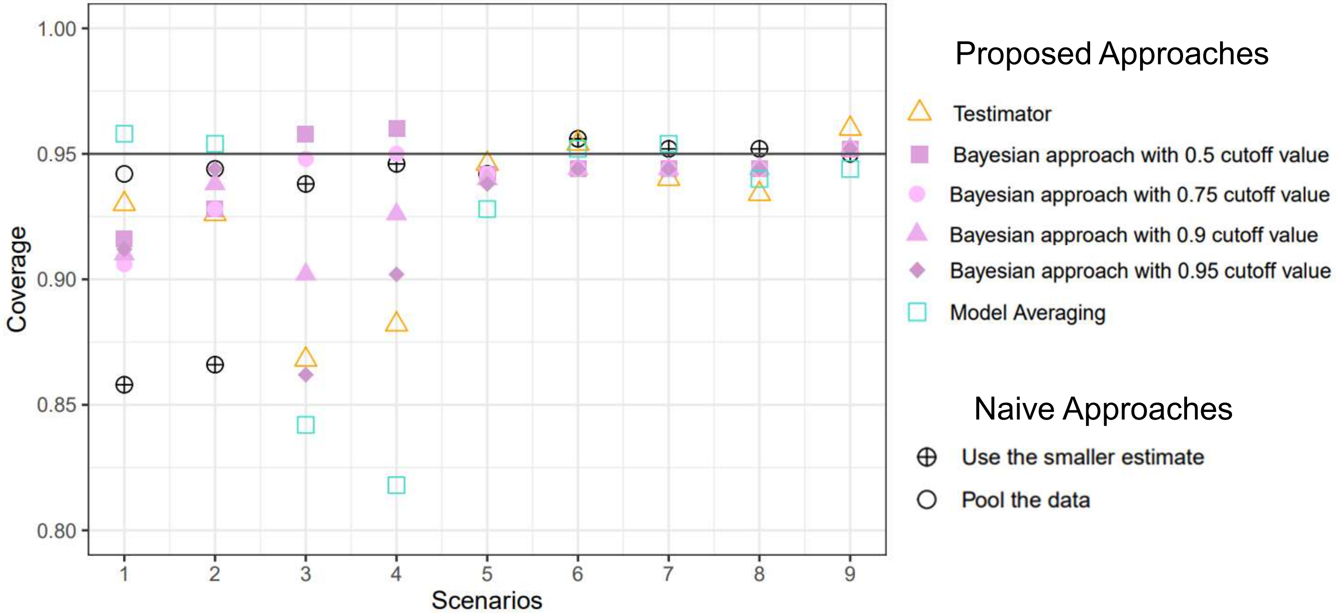
Simulation results（prior info available）



NO	Scenarios	Effect size
1	$u_b = 0, \sigma_b^2 = 1$	0
2	$u_b = 0, \sigma_b^2 = 2$	0
3	$u_b = 0.3, \sigma_b^2 = 2$	0.24
4	$u_b = 0.3\ \sigma_b^2 = 1$	0.30
5	$u_b = 0.5, \sigma_b^2 = 2$	0.41
6	$u_b = 0.5\ \sigma_b^2 = 1$	0.50
7	$u_b = 0.7, \sigma_b^2 = 2$	0.57
8	$u_b = 0.7, \sigma_b^2 = 1$	0.70
9	$u_b = 0.7, \sigma_b^2 = 0.5$	0.81

When no mode effects, the proposed approaches reduce bias, compared to taking the smaller estimate
When mode effects are present, the proposed approaches lessen bias, compared to pooling the data

Simulation results（prior info available）



NO	Scenarios	Effect size
1	$u_b = 0, \sigma_b^2 = 1$	0
2	$u_b = 0, \sigma_b^2 = 2$	0
3	$u_b = 0.3, \sigma_b^2 = 2$	0.24
4	$u_b = 0.3\ \sigma_b^2 = 1$	0.30
5	$u_b = 0.5, \sigma_b^2 = 2$	0.41
6	$u_b = 0.5\ \sigma_b^2 = 1$	0.50
7	$u_b = 0.7, \sigma_b^2 = 2$	0.57
8	$u_b = 0.7, \sigma_b^2 = 1$	0.70
9	$u_b = 0.7, \sigma_b^2 = 0.5$	0.81

In scenarios 3-9, the Naïve approach of pooling the data has a nearly 0 coverage rate (not shown)
When no mode effects, the proposed approaches have better coverage than taking the smaller estimate
When mode effects are present, the proposed approaches achieve better coverage than pooling the data

Application: Jordan experiment

Outcome: satisfaction with the government (sum of six 4-point items, range: 6-24).

- The higher the measure, the more satisfied a participant is with the government
- Reporting low satisfaction can be risky; thus, we consider a smaller estimate to be less biased

Randomized mode assignment

- Compute nonresponse weights to account for nonresponse to the telephone interview

Complex sample design

- Stratified by governorates
- 1) Sample blocks with proportional to size sampling and 2) sample households with SRSWOR
- Final weights = Base weights * Calibration weights * Nonresponse weights

Compute design-based statistics for each mode and plug them into the three approaches

- Sample means, sample variances, and effective sample sizes

Consider a smaller estimate is less biased

	Proposed approaches			Naïve approaches		
	Testimator	Bayesian	Model Averaging	Pool data	FTF alone	TEL alone
Point Est	7.737	7.738	7.739	8.479	7.737	9.154
95% Interval	7.443, 8.031	7.433, 8.036	7.454, 8.020	8.266, 8.691	7.437, 8.038	8.916, 9.392

The three approaches provide similar point estimates as using FTF data alone

Conclusion

1. When having prior info about preferred directions, the approaches provide robust inferences
2. When no info, the approaches provide intervals with nominal coverage rates (not shown)
3. The three approaches can be adapted to account for complex sample designs

Future work

1. Account for non-randomized mixed-mode designs using propensity score adjustments
2. Incorporate the prior information about preferred modes
3. Extend the approaches to binary outcomes
4. In the Testimator approach, find optimal cutoff values(α) to minimize MSE