Robust Inference for Audit Studies

April 15, 2021

Daniel S. Putman, Postdoctoral Fellow

Mystery Shopping with Agents | Innovations for Poverty Action

Motivation: Robust Inference for Audit Studies

1. Widely used to measure differential outcomes in markets
   - Bertrand and Mullainathan (2004): labor outcomes by race
   - Annan (2020): mobile money overcharging by gender

2. False positives about discrimination can come with heavy consequences (e.g., regulatory attention)

Research Questions

1. What are appropriate standard errors?
   - Abadie et al. (2017): study design and sampling design should guide our choice
   - However, audit studies provide many different rationales
2. How much does getting it wrong matter?
3. Can we provide tools for design of powered studies that account for robust inference?

A Mobile Money Overcharging Example

Research question: Are women charged more than men when making cash out transactions at mobile money agents?

Sampling design: sample agents

Study design: we assign male and female shoppers to complete cash out transactions at multiple agents

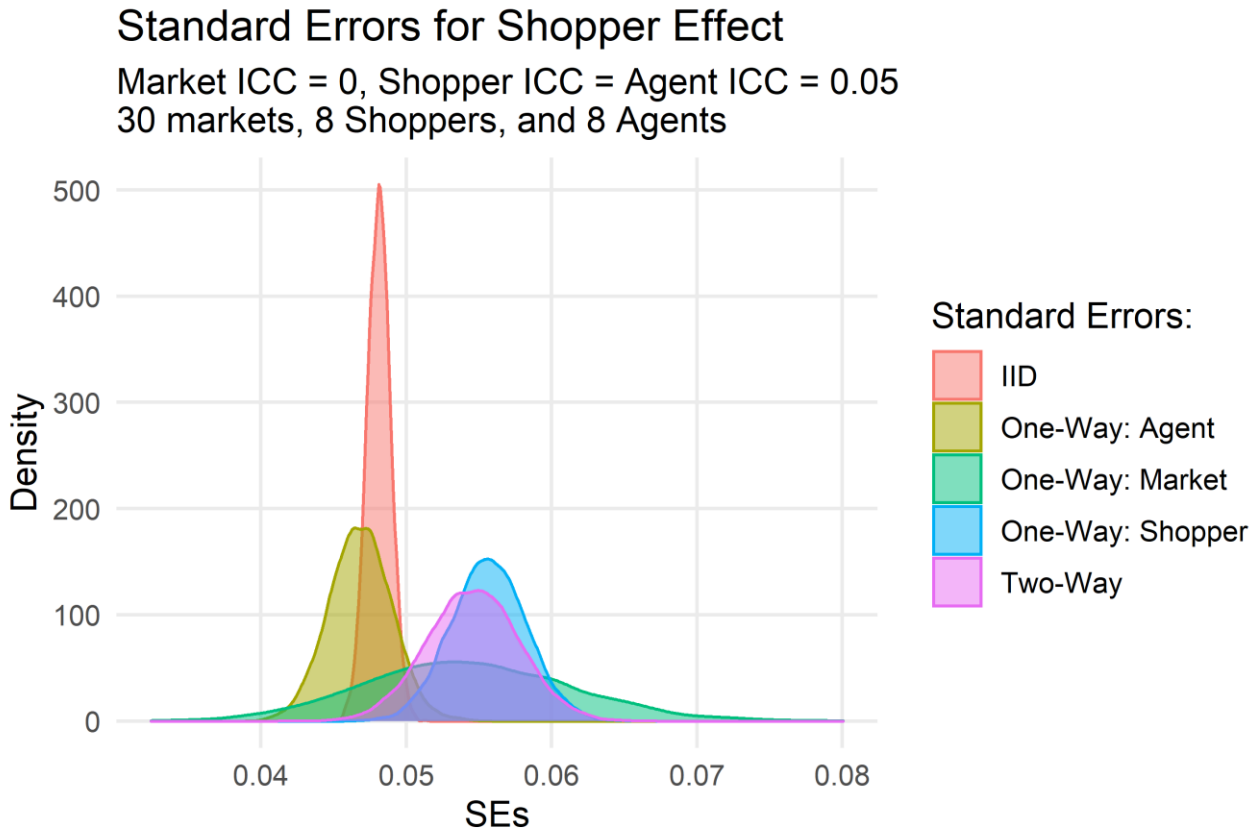
Analysis: We regress the charges on gender to measure the differences in outcome by gender (β below)

$$Charge_{ij} = \alpha + \beta\ Female_i + \varepsilon_{ij}$$

Here we would want cluster on agent (sampling) or shopper (study design), or both

How much does it matter?

- Clustering one-way on shopper and two-way clustering are balanced in terms of false positives
- Clustering one-way on agent is more liberal than i.i.d. errors
- Clustering on market is also liberal
  - Too few clusters (Cameron and Miller, 2015)
  - Wild cluster bootstrap (MacKinnon and Webb, 2018)



Power and Audit Study Design

- Two-way clustering isn't an option in standard commands
- Burlig et al. (2020) offers simulation and analytical calculations for panel RCTs
  - One-way clustering and serial correlation
  - Can we build something similar for audit studies?
- Training auditors is difficult, often working with few auditors per market and few markets – how to trade off markets, auditors, and agents?

References

Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering?

Annan, F. (2020). Gender and Financial Misconduct on Fintech: Experimental Evidence from Mobile Money in Ghana. *SSRN Electronic Journal*.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*.

Burlig, F., Preonas, L., & Woerman, M. (2020). Panel data and experimental design. *Journal of Development Economics*, *144*(February), 102458.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372.

MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *Econometrics Journal*, *21*(2), 114–135.

Data Generating Process: Simulations

Data is generated:

$$y_{ijm} = \mu + \beta T_{im} + \gamma_{im} + \delta_{jm} + \varepsilon_{ijm}$$

- $T_i = 1$ shopper i is female, 0 male
- Simulate under the null (so $\beta = 0$), ignore shifts in outcome ($\mu = 0$)
- ρ_S is shopper ICC, ρ_A is agent ICC, I use $\rho_S = \rho_A = 0.05$
- Idiosyncratic shock $\varepsilon_{ijm} \sim N(0, \sigma_\varepsilon^2)$. I use $\sigma_\varepsilon^2 = 1$
- Shopper shock: $\gamma_{im} \sim N(0, \sigma_\gamma^2)$, where $\sigma_\gamma^2 = \frac{\rho_S \sigma_\varepsilon^2}{1-\rho_S-\rho_A}$
- Agent shock: $\gamma_{jm} \sim N(0, \sigma_\delta^2)$, where $\sigma_\delta^2 = \frac{\rho_A \sigma_\varepsilon^2}{1-\rho_S-\rho_A}$

More Results: False Positive Simulations

			Rejection Rate under Null			
			Gender of	Factorial Model: Gender of		
Standard Errors	n_S	n_A	Shopper	Shopper	Agent	Interaction
IID	16	4	6.02	5.08	8.52	3.94
	8	8	8.76	6.44	6.18	3.40
	4	16	12.76	7.98	4.70	3.18
One-Way: Agent	16	4	6.90	6.20	5.36	4.84
	8	8	9.80	7.30	4.84	4.18
	4	16	13.80	9.02	4.36	3.82
One-Way: Shopper	16	4	4.46	4.60	9.46	4.66
	8	8	4.76	4.88	6.90	4.12
	4	16	5.08	4.32	5.82	3.94
Two-Way: Shopper & Agent	16	4	5.12	5.70	5.82	5.44
	8	8	5.08	5.56	5.40	5.10
	4	16	5.40	4.64	5.24	4.66
One-Way: Market	16	4	5.62	5.80	6.26	5.98
	8	8	5.96	6.02	5.88	5.74
	4	16	5.76	5.26	5.72	5.32

Table 2: False Positives in Audit Studies by Choice of Standard Error