Lying on Surveys

Methods for List Experiments with Direct Questioning

Winston Chou

Princeton University

February 15, 2019

Introduction

- Survey methodology for sensitive questions

Introduction

- Survey methodology for sensitive questions
- Elicitation of traits whose revelation is potentially harmful to respondents

Introduction

- Survey methodology for sensitive questions
- Elicitation of traits whose revelation is potentially harmful to respondents
    - Unpopular opinions or uncomfortable facts

Introduction

- Survey methodology for sensitive questions
- Elicitation of traits whose revelation is potentially harmful to respondents
  - Unpopular opinions or uncomfortable facts
  - Health data, incarceration history

Introduction

- Survey methodology for sensitive questions
- Elicitation of traits whose revelation is potentially harmful to respondents
  - Unpopular opinions or uncomfortable facts
  - Health data, incarceration history
  - Support for combatants in wartime (Lyall, Blair, and Imai 2013; Matanock and Garcia-Sanchez 2018)

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms
  - Desire to keep face or project certain identity

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms
  - Desire to keep face or project certain identity
- Lack of privacy or plausible deniability

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms
  - Desire to keep face or project certain identity
- Lack of privacy or plausible deniability
- ...whether at point of collection or in event of deanonymization

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms
  - Desire to keep face or project certain identity
- Lack of privacy or plausible deniability
- ... whether at point of collection or in event of deanonymization

Direct Questioning

- Direct questioning can be notoriously unreliable for sensitive questions.
- Why?
  - Fear of retaliation
  - Social norms
  - Desire to keep face or project certain identity
- Lack of privacy or plausible deniability
- ...whether at point of collection or in event of deanonymization

Sensitive Trait		**Direct Response**
$Z_i^* = 0$	$\longleftarrow$	$D_i = 0$
$Z_i^* = 1$	$\longleftarrow$	$D_i = 1$

Figure: Identifiability Under Direct Questioning

Indirect Questioning

- Break 1:1 mapping by deliberately corrupting signal

Indirect Questioning

- Break 1:1 mapping by deliberately corrupting signal
- Afford respondents privacy ($\uparrow$ variance)

Indirect Questioning

- Break 1:1 mapping by deliberately corrupting signal
- Afford respondents privacy (↑ variance)
- Encourage respondents to provide truthful answers (↓ bias)

Randomized Response Technique (Warner 1965)

- Indirect questioning technique that obscures responses with random noise.

Randomized Response Technique (Warner 1965)

- Indirect questioning technique that obscures responses with random noise.
- Respondent uses a randomization device (e.g., a quarter) whose distribution but not outcome is known to the enumerator:

Randomized Response Technique (Warner 1965)

- Indirect questioning technique that obscures responses with random noise.
- Respondent uses a randomization device (e.g., a quarter) whose distribution but not outcome is known to the enumerator:

Randomized Response Technique (Warner 1965)

- Indirect questioning technique that obscures responses with random noise.
- Respondent uses a randomization device (e.g., a quarter) whose distribution but not outcome is known to the enumerator:

```
Flip this coin, but do not tell me whether it landed heads or
tails.  If it landed heads, please answer the following question
truthfully.  If it landed tails, please say "YES."

Many people fail to vote for one reason or another.  What about
you?  Did you fail to vote?
```

Sensitive Trait **Indirect Response**

$$
\begin{tikzcd}
Z_i^* = 0 & & Y_i = 0 \arrow[ll] \\
Z_i^* = 1 & & Y_i = 1 \arrow[ll] \arrow[ull]
\end{tikzcd}
$$

Figure: Privacy Under Indirect Questioning

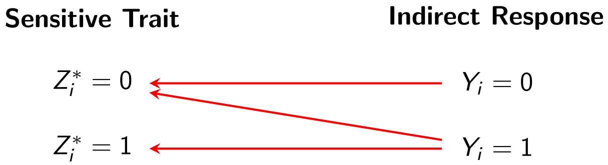


Figure: Privacy Under Indirect Questioning

Letting $\pi = \mathbb{E}[Z_i^*]$,

$$\prod_{i=1}^{N} \left(\frac{1}{2} + \frac{1}{2}\pi\right)^{\mathbf{1}(Y_i=1)} \left[\frac{1}{2}(1-\pi)\right]^{\mathbf{1}(Y_i=0)}. \tag{1}$$

List Experiment

- Indirect questioning technique

List Experiment

- Indirect questioning technique
- Experiment where the "treatment effect" is unbiased (under assumptions) for the prevalence of the sensitive trait ($\mathbb{E}[Z_i^*]$)

Coffman, Coffman, and Ericson (2016)

Control Condition

- I have more than one bathroom in my home.
- I can walk to a grocery store from my home in less than 30 minutes.
- I can speak more than one language.
- All four of my grandparents were born in the United States.

Please fill in the bubble that corresponds to the total number of statements that apply to you (0-4).

Coffman, Coffman, and Ericson (2016)

Treatment Condition

- I have more than one bathroom in my home.
- I can walk to a grocery store from my home in less than 30 minutes.
- I can speak more than one language.
- All four of my grandparents were born in the United States.
- I would be unhappy to have an openly lesbian, gay, or bisexual manager at work.

Please fill in the bubble that corresponds to the total number of statements that apply to you (0, 5).

$(Z_i^* = 0, Y_{0i} = 0)$ $(Z_i^* = 1, Y_{0i} = 0)$

$(Z_i^* = 0, Y_{0i} = 1)$ $(Z_i^* = 1, Y_{0i} = 1)$

$(Z_i^* = 0, Y_{0i} = 2)$ $(Z_i^* = 1, Y_{0i} = 2)$

$(Z_i^* = 0, Y_{0i} = 3)$ $(Z_i^* = 1, Y_{0i} = 3)$

$(Z_i^* = 0, Y_{0i} = 4)$ $(Z_i^* = 1, Y_{0i} = 4)$

$$(Z_i^* = 0, Y_{0i} = 0)$$

$$(Z_i^* = 0, Y_{0i} = 1), (Z_i^* = 1, Y_{0i} = 0)$$

$$(Z_i^* = 0, Y_{0i} = 2), (Z_i^* = 1, Y_{0i} = 1)$$

$$(Z_i^* = 0, Y_{0i} = 3), (Z_i^* = 1, Y_{0i} = 2)$$

$$(Z_i^* = 0, Y_{0i} = 4), (Z_i^* = 1, Y_{0i} = 3)$$

$$(Z_i^* = 1, Y_{0i} = 4)$$

Truth		**Indirect Response**
$(Z_i^* = 0, Y_{0i} = 0)$	$\longrightarrow$	$Y_{1i} = 0$
$(Z_i^* = 0, Y_{0i} = 1)$, $(Z_i^* = 1, Y_{0i} = 0)$	$\longrightarrow$	$Y_{1i} = 1$
$(Z_i^* = 0, Y_{0i} = 2)$, $(Z_i^* = 1, Y_{0i} = 1)$	$\longrightarrow$	$Y_{1i} = 2$
$(Z_i^* = 0, Y_{0i} = 3)$, $(Z_i^* = 1, Y_{0i} = 2)$	$\longrightarrow$	$Y_{1i} = 3$
$(Z_i^* = 0, Y_{0i} = 4)$, $(Z_i^* = 1, Y_{0i} = 3)$	$\longrightarrow$	$Y_{1i} = 4$
$(Z_i^* = 1, Y_{0i} = 4)$	$\longrightarrow$	$Y_{1i} = 5$

Figure: List Experiment Structure

Truth	**Indirect Response**
$(Z_i^* = 0, Y_{0i} = 0)$	$Y_{1i} = 0$
$(Z_i^* = 0, Y_{0i} = 1)$, $(Z_i^* = 1, Y_{0i} = 0)$	$Y_{1i} = 1$
$(Z_i^* = 0, Y_{0i} = 2)$, $(Z_i^* = 1, Y_{0i} = 1)$	$Y_{1i} = 2$
$(Z_i^* = 0, Y_{0i} = 3)$, $(Z_i^* = 1, Y_{0i} = 2)$	$Y_{1i} = 3$
$(Z_i^* = 0, Y_{0i} = 4)$, $(Z_i^* = 1, Y_{0i} = 3)$	$Y_{1i} = 4$
$(Z_i^* = 1, Y_{0i} = 4)$	$Y_{1i} = 5$

Figure: List Experiment

Statistical Analysis of the List Experiment

Identification (Blair and Imai 2012)

- Ignorable treatment assignment

Statistical Analysis of the List Experiment

Identification (Blair and Imai 2012)

- Ignorable treatment assignment
- No design effects

$$\begin{aligned} Y_{ij}(0) &= Y_{ij}(1) \text{ for } j = 1, \ldots J & (2) \\ Y_i(T_i) &= \sum_{j=1}^{J} Y_{ji} + T_i Z_i(T_i). & (3) \end{aligned}$$

- Ignorable treatment assignment
- No design effects

$$\begin{aligned} Y_{ij}(0) &= Y_{ij}(1) \text{ for } j = 1, \dots J & (2) \\ Y_i(T_i) &= \sum_{j=1}^{J} Y_{ji} + T_i Z_i(T_i). & (3) \end{aligned}$$

- No liars

$$Z_i(1) = Z_i^*. \quad (4)$$

This Paper

- "Lying on Surveys"

This Paper

- "Lying on Surveys"
- Beyond measuring sensitive traits $\rightsquigarrow$ Who misreports?

This Paper

- "Lying on Surveys"
- Beyond measuring sensitive traits $\rightsquigarrow$ Who misreports?
  - Ex. Are Democrats or Republicans more likely to misreport homophobic attitudes? (Coffman, Coffman, and Ericson 2016)

This Paper

- "Lying on Surveys"
- Beyond measuring sensitive traits $\rightsquigarrow$ Who misreports?
  - Ex. Are Democrats or Republicans more likely to misreport homophobic attitudes? (Coffman, Coffman, and Ericson 2016)
  - Ex. Does territorial control predict preference falsification? (Matanock and Garcia-Sanchez 2018)

This Paper

- "Lying on Surveys"
- Beyond measuring sensitive traits $\rightsquigarrow$ Who misreports?
  - Ex. Are Democrats or Republicans more likely to misreport homophobic attitudes? (Coffman, Coffman, and Ericson 2016)
  - Ex. Does territorial control predict preference falsification? (Matanock and Garcia-Sanchez 2018)
- Combine list experiments with direct questioning (Eady 2017)

Misreporting Rate

Nonparametric Identification

We want to estimate $\Pr(D_i = 0 | Z_i^* = 1)$

Misreporting Rate
Nonparametric Identification

We want to estimate $\Pr(D_i = 0 | Z_i^* = 1)$

Assume **monotonicity**:

$$\Pr(D_i = 1 | Z_i^* = 0) = 0. \tag{5}$$

Misreporting Rate

Nonparametric Identification

We want to estimate $\Pr(D_i = 0|Z_i^* = 1)$

Assume **monotonicity**:

$$\Pr(D_i = 1|Z_i^* = 0) = 0. \tag{5}$$

By Bayes's Rule:

$$Pr(D_i = 0|Z_i^* = 1) \quad = \quad 1 - Pr(Z_i^* = 1|D_i = 1)Pr(D_i = 1)/Pr(Z_i^* = 1). \tag{6}$$

- $\Pr(Z_i^* = 1)$ given by list experiment
- $\Pr(Z_i^* = 1|D_i = 1) = 1$ by monotonicity
- $\Pr(D_i = 1)$ given by direct questioning.

Ratio Estimator

$$\rightsquigarrow 1 - \frac{Pr(D_i = 1)}{Pr(Z_i^* = 1)} \tag{7}$$

Ratio Estimator

$$1 - \frac{Pr(D_i = 1)}{Pr(Z_i^* = 1)} \tag{8}$$

$\mathbb{E}[D_i]$, $\mathbb{E}[Z_i^*]$ can be from different surveys.

Regression Modeling

- How does misreporting vary in the population?

Regression Modeling

- How does misreporting vary in the population?

Maximum Likelihood Model for List Experiments (Imai 2011, Blair and Imai 2012, `list`)

$$
\begin{aligned}
Y_i(0) | X_i &\sim \text{Binomial}(J, f(X_i; \gamma)) && (9) \\
Z_i^* | X_i &\sim \text{Bernoulli}(g(X_i; \delta)). && (10)
\end{aligned}
$$

Regression Modeling

- How does misreporting vary in the population?

Maximum Likelihood Model for List Experiments (Imai 2011, Blair and Imai 2012, `list`)

$$
\begin{aligned}
Y_i(0)|X_i &\sim \text{Binomial}(J, f(X_i;\gamma)) && (9)\\
Z_i^*|X_i &\sim \text{Bernoulli}(g(X_i;\delta)). && (10)
\end{aligned}
$$

Model for Misreporting (Eady 2017)

$$
D_i|Z_i^* = 1, X_i \sim \text{Bernoulli}(1 - h(X_i;\beta)). \quad (11)
$$

Maximum Likelihood

- For simplicity, suppose direct question is only posed to control respondents (partially-overlapping design).

Maximum Likelihood

- For simplicity, suppose direct question is only posed to control respondents (partially-overlapping design).

Control Observations

$$
\begin{aligned}
\Pr(D_i = 1) &= g(X_i; \delta)[1 - h(X_i; \beta)] && (12) \\
\Pr(D_i = 0) &= 1 - g(X_i; \delta) + g(X_i; \delta) h(X_i; \beta) && (13) \\
\Pr(Y_i = y) &= \text{Bin}(y; J, f(X_i; \gamma)). && (14)
\end{aligned}
$$

Maximum Likelihood

- For simplicity, suppose direct question is only posed to control respondents (partially-overlapping design).

Control Observations

$$\begin{aligned}
\Pr(D_i = 1) &= g(X_i;\delta)[1 - h(X_i;\beta)] && (12)\\
\Pr(D_i = 0) &= 1 - g(X_i;\delta) + g(X_i;\delta)h(X_i;\beta) && (13)\\
\Pr(Y_i = y) &= \mathrm{Bin}(y; J, f(X_i;\gamma)). && (14)
\end{aligned}$$

Treatment Observations

$$\begin{aligned}
\Pr(Y_i = 0) &= \mathrm{Bin}(0; J, f(X_i;\gamma))[1 - g(X_i;\delta)] && (15)\\
\Pr(Y_i = y) &= g(X_i;\delta)\mathrm{Bin}(y-1; J, f(X_i;\gamma)) + && (16)\\
&\qquad [1 - g(X_i;\delta)]\mathrm{Bin}(y; J, f(X_i;\gamma)) &&\\
\Pr(Y_i = J+1) &= \mathrm{Bin}(J; J, f(X_i;\gamma))g(X_i;\delta). && (17)
\end{aligned}$$

EM Algorithm

EM Algorithm

Control Observations

$$\Pr(D_i = 0) \quad = \quad 1 - g(X_i; \delta) + g(X_i; \delta) h(X_i; \beta). \tag{18}$$

Treatment Observations

$$\begin{aligned} \Pr(Y_i = y) \quad = \quad & g(X_i; \delta) \text{Bin}(y - 1; J, f(X_i; \gamma)) + \\ & [1 - g(X_i; \delta)] \text{Bin}(y; J, f(X_i; \gamma)). \end{aligned} \tag{19}$$

0. Write down the log-likelihood we wish we could solve, replacing latent variables with expectations (Q-function).
1. Update the expectations given the current parameter estimates (E-step).
2. Given the expectations, maximize to update the parameter estimates (M-step).

Iterate steps 1 and 2 until convergence.

EM Algorithm

Supposing we observe Z_i^*:

Control Observations

$$\begin{aligned}
\Pr(D_i = 0, Z_i^* = 1) &= g(X_i; \delta) h(X_i; \beta) && (20) \\
\Pr(D_i = 0, Z_i^* = 0) &= 1 - g(X_i; \delta). && (21)
\end{aligned}$$

Treatment Observations

$$\begin{aligned}
\Pr(Y_i(1) = y, Z_i^* = 1) &= g(X_i; \delta) \text{Bin}(y - 1; J, f(X_i; \gamma)) && (22) \\
\Pr(Y_i(1) = y, Z_i^* = 0) &= [1 - g(X_i; \delta)] \text{Bin}(y; J, f(X_i; \gamma)). && (23)
\end{aligned}$$

In the E-step, we just compute $\mathbb{E}[\mathbf{1}(Z_i^* = z)] = \frac{a}{a+b}$, $z = 0, 1$.

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).
- Large differences-in-differences observed for Christians (+12%) versus nonreligious and Republicans (+35%) versus Democrats

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).
- Large differences-in-differences observed for Christians (+12%) versus nonreligious and Republicans (+35%) versus Democrats
- But these variables are strongly correlated

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).
- Large differences-in-differences observed for Christians (+12%) versus nonreligious and Republicans (+35%) versus Democrats
- But these variables are strongly correlated
  - 70% of R respondents identify as Christian, compared to 30% of Ds

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).
- Large differences-in-differences observed for Christians (+12%) versus nonreligious and Republicans (+35%) versus Democrats
- But these variables are strongly correlated
    - 70% of R respondents identify as Christian, compared to 30% of Ds
    - Christianity strongly correlated with religiosity ($\rho = 0.60$)

Application: Homophobia in the Workplace (CCE 2016)

"I would be unhappy to have an openly lesbian, gay, or bisexual manager at work."

- Partially-overlapping design, only control respondents answer DQ
- Respondents 67% more likely to express discomfort with openly gay manager at work under list experiment (27%) versus direct questioning (16%).
- Large differences-in-differences observed for Christians (+12%) versus nonreligious and Republicans (+35%) versus Democrats
- But these variables are strongly correlated
  - 70% of R respondents identify as Christian, compared to 30% of Ds
  - Christianity strongly correlated with religiosity ($\rho = 0.60$)
  - Rs score about 1 standard deviation higher on religiosity scale

Multiple Regression

- How does discomfort with LGB managers vary? $g(X_i; \delta)$

Multiple Regression

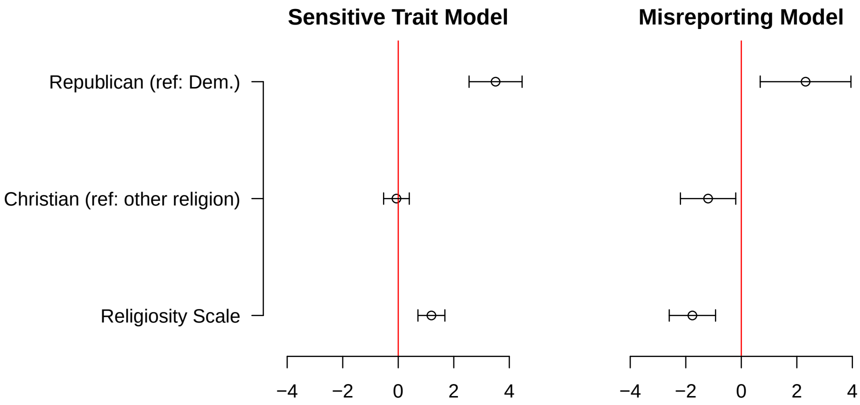
- How does discomfort with LGB managers vary? $g(X_i; \delta)$
- How does the propensity to underreport this vary? $h(X_i; \beta)$

Multiple Regression

- How does discomfort with LGB managers vary? $g(X_i;\delta)$
- How does the propensity to underreport this vary? $h(X_i;\beta)$

Multiple Regression

- How does discomfort with LGB managers vary? $g(X_i; \delta)$
- How does the propensity to underreport this vary? $h(X_i; \beta)$



Leveraging Floor and Ceiling Effects

$$\begin{array}{ccc} \textbf{Sensitive Trait} & & \textbf{Direct Response} \\ Z_i^* = 0 & \longleftarrow & D_i = 0 \\ Z_i^* = 1 & \longleftarrow & D_i = 1 \end{array}$$

Figure: Identifiability Under Direct Questioning

- Floor and ceiling responses are analogous to direct questioning

Leveraging Floor and Ceiling Effects

Sensitive Trait		**Direct Response**
$Z_i^* = 0$	$\longleftarrow$	$D_i = 0$
$Z_i^* = 1$	$\longleftarrow$	$D_i = 1$

Figure: Identifiability Under Direct Questioning

- Floor and ceiling responses are analogous to direct questioning
- May also be subject to misreporting and nonresponse (Glynn 2013)

Leveraging Floor and Ceiling Effects

Sensitive Trait		**Direct Response**
$Z_i^* = 0$	$\longleftarrow$	$D_i = 0$
$Z_i^* = 1$	$\longleftarrow$	$D_i = 1$

Figure: Identifiability Under Direct Questioning

- Floor and ceiling responses are analogous to direct questioning
- May also be subject to misreporting and nonresponse (Glynn 2013)
- Ex.: Support for Taliban (Lyall, Blair, and Imai 2013)

Lyall, Blair, and Imai (2013)

[...] I'd like you to tell me how many of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don't tell me which ones you generally agree with; only tell me how many groups or individuals you broadly support.
Karzai Government; National Solidarity Program; Local Farmers,
Taliban

Lyall, Blair, and Imai (2013)

`[...] I'd like you to tell me how many of these groups and individuals you broadly support, meaning that you generally agree with the goals and policies of the group or individual. Please don't tell me which ones you generally agree with; only tell me how many groups or individuals you broadly support.`
`Karzai Government; National Solidarity Program; Local Farmers, Taliban`

	Control group		Treatment group	
response	counts	percentage	counts	percentage
0	188	20	0	0
1	265	29	433	47
2	265	29	287	31
3	200	22	198	22
4			0	0

Truth		Indirect Response
$(Z_i^* = 0, Y_{0i} = 0)$	$\longleftarrow$	$Y_{1i} = 0$
$(Z_i^* = 0, Y_{0i} = 1)$, $(Z_i^* = 1, Y_{0i} = 0)$	$\dashleftarrow$	$Y_{1i} = 1$
$(Z_i^* = 0, Y_{0i} = 2)$, $(Z_i^* = 1, Y_{0i} = 1)$	$\dashleftarrow$	$Y_{1i} = 2$
$(Z_i^* = 0, Y_{0i} = 3)$, $(Z_i^* = 1, Y_{0i} = 2)$	$\dashleftarrow$	$Y_{1i} = 3$
$(Z_i^* = 0, Y_{0i} = 4)$, $(Z_i^* = 1, Y_{0i} = 3)$	$\dashleftarrow$	$Y_{1i} = 4$
$(Z_i^* = 1, Y_{0i} = 4)$	$\longleftarrow$	$Y_{1i} = 5$

Figure: Floor and Ceiling Responses Analogous to Direct Questioning

Ceiling Effects Model

- Treatment respondents can be informative about misreporting, even in partially-overlapping designs

Ceiling Effects Model

- Treatment respondents can be informative about misreporting, even in partially-overlapping designs

$$\begin{aligned}
\Pr(Y_i(1) = J+1) &= g(X_i;\delta)[1-h(X_i;\beta)]f(X_i;\gamma)^J && (24)\\
\Pr(Y_i(1) = J) &= g(X_i;\delta)h(X_i;\beta)f(X_i;\gamma)^J + && (25)\\
&\quad g(X_i;\delta)Jf(X_i;\gamma)^{J-1}[1-f(X_i;\gamma)] + \\
&\quad [1-g(X_i;\delta)]f(X_i;\gamma)^J.
\end{aligned}$$

- Potential efficiency gains (smaller standard errors)

- The monotonicity assumption is that the sensitive response is the same for everyone.

- The monotonicity assumption is that the sensitive response is the same for everyone.
- Either the underreporting rate equals zero or overreporting rate equals zero

Two-Way Misreporting

Motivation

- The monotonicity assumption is that the sensitive response is the same for everyone.
- Either the underreporting rate equals zero or overreporting rate equals zero
- This assumption may be unrealistic depending on the sampling frame

Two-Way Misreporting
Motivation

- The monotonicity assumption is that the sensitive response is the same for everyone.
- Either the underreporting rate equals zero or overreporting rate equals zero
- This assumption may be unrealistic depending on the sampling frame
- In the fully overlapping design, both underreporting and overreporting are nonparametrically identified.

Two-Way Misreporting

Nonparametric Identification

In fully-overlapping designs, we don't actually need monotonicity to identify $Pr(D_i = 0|Z_i^* = 1)$:

$$
\begin{aligned}
& E[Y_i|D_i(1) = 1, T_i = 1] - E[Y_i|D_i(0) = 1, T_i = 0] && (26) \\
= \; & E[Y_i(1) - Y_i(0)|D_i = 1] && (27) \\
= \; & \Pr(Z_i^* = 1|D_i = 1) && (28) \\
= \; & \frac{\Pr(D_i = 1|Z_i^* = 1)\Pr(Z_i^* = 1)}{\Pr(D_i = 1)}. && (29)
\end{aligned}
$$

- Need $D_i(1) = D_i(0) = 1$ (direct response unaffected by treatment)
- By same argument, $Pr(D_i = 1|Z_i^* = 0)$ is also identified.

Two-Way Misreporting

Nonparametric Identification

In fully-overlapping designs, we don't actually need monotonicity to identify $Pr(D_i = 0|Z_i^* = 1)$:

$$
\begin{aligned}
& E[Y_i|D_i(1) = 1, T_i = 1] - E[Y_i|D_i(0) = 1, T_i = 0] && (30) \\
= \quad & E[Y_i(1) - Y_i(0)|D_i = 1] && (31) \\
= \quad & \Pr(Z_i^* = 1|D_i = 1) && (32) \\
= \quad & \frac{\Pr(D_i = 1|Z_i^* = 1)\Pr(Z_i^* = 1)}{\Pr(D_i = 1)}. && (33)
\end{aligned}
$$

- Need $D_i(1) = D_i(0) = 1$ (direct response unaffected by treatment)
- By same argument, $Pr(Z_i^* = 1|D_i = 0)$ is also identified.
- In the fully overlapping design, both underreporting and overreporting are nonparametrically identified

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse
- Statistical analysis of the list experiment is tricky

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse
- Statistical analysis of the list experiment is tricky
  - Bias-variance tradeoff can be steep (Blair, Coppock, and Moor n.d.)

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse
- Statistical analysis of the list experiment is tricky
  - Bias-variance tradeoff can be steep (Blair, Coppock, and Moor n.d.)
  - Low variance models can be sensitive to assumptions (Ahlquist 2018, Blair, Chou, and Imai forthcoming)

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse
- Statistical analysis of the list experiment is tricky
  - Bias-variance tradeoff can be steep (Blair, Coppock, and Moor n.d.)
  - Low variance models can be sensitive to assumptions (Ahlquist 2018, Blair, Chou, and Imai forthcoming)
- Combining list experiments with direct questioning can improve aspects of both (Aronow et al. 2017)

Concluding Remarks

- List experiment can be useful for reducing bias from misreporting and nonresponse
- Statistical analysis of the list experiment is tricky
  - Bias-variance tradeoff can be steep (Blair, Coppock, and Moor n.d.)
  - Low variance models can be sensitive to assumptions (Ahlquist 2018, Blair, Chou, and Imai forthcoming)
- Combining list experiments with direct questioning can improve aspects of both (Aronow et al. 2017)
- Misreporting models as a side benefit (Eady 2017)

Further Reading

List Experiments

- Imai 2011 JASA on identification and regression modeling
- Glynn 2013 POQ on floor and ceiling effects and selection of control items

Statistical Analysis

- Ahlquist 2018 PA on sensitivity of ML regression model
- Blair, Chou, and Imai forthcoming PA on robust ML models
  - Here we find that prevalence is the key issue
  - ML generally agrees with NLS unless sensitive item is somewhat rare ($\leq 15\%$)

Misreporting

- Eady 2017 PA on combining list experiments and direct questioning
- My working paper (`http://princeton.edu/~wchou`)
  - Extending framework to partially-overlapping survey designs
  - Leveraging floor and ceiling effects for better inference
  - Two-way misreporting identification and regression