Improving Survey Methodologies for Sensitive Questions

Kosuke Imai

Princeton University

July 5, 2017
Seminar Talk at the University of Tokyo

Joint work with
Graeme Blair (UCLA), Jason Lyall (Yale), Bryn Rosenfeld (USC),
and Winston Chou (Princeton)

Introduction

- Challenge of measuring sensitive attitudes and behaviors
    - social desirability bias
    - non-response bias
- Racial prejudice, corruption, support for controversial political actors
- Indirect methods becoming increasingly popular
    - list experiments: aggregation
    - endorsement experiment: priming
    - randomized response: randomization
- Development of statistical methods
    - multivariate regression for each survey technique (Bullock, Imai and Shapiro 2010; Imai 2011; Blair & Imai 2012; Blair, Imai & Zhou 2015)
    - using responses as predictors in outcome regression (Imai, Park & Greene 2015)
- Empirical validation studies
    - validation against ground truth (Rosenfeld, Imai & Shapiro 2016)
    - prediction of behavior (Hirose, Imai & Lyall 2017)

Two Ways to Improve Sensitive Question Survey Methods

1. Comparing and combining multiple measurements (Blair, Imai & Lyall 2014)
   - Agreement among multiple measurements ⤳ more credible
   - Combining multiple measurements ⤳ more powerful
   - Application: Hearts and minds in Afghanistan

2. Using auxiliary information (Chou, Imai, & Rosenfeld 2017)
   - Sometimes aggregate truths are available
     - Turnout rates and voting outcomes
     - Administrative records, e.g., crime and incarceration
   - Use auxiliary information to improve individual-level inference
   - Application: Missisippi anti-abortion referendum

Hearts and Minds in Afghanistan (Blair, Imai & Lyall, 2014)

► How do we measure civilian attitudes in a conflict setting?

► Current efforts in Afghanistan rely on direct questions:
    1. USAID (TCAPF): "Who do you believe can solve your problems?"
    2. ISAF (ANQAR): "Over the past 6 months, do you think the Taliban have grown stronger, grown weaker, or remained the same?"

► Why are direct questions a bad idea?
    1. Threats to enumerators and respondents
    2. Nonresponse, social desirability bias
    3. Interviews are public
    4. Danger of selection bias in sampling locations (role of gatekeepers)

► ANQAR (November-December 2011): 50% refusal rate

Public Nature of Interviews
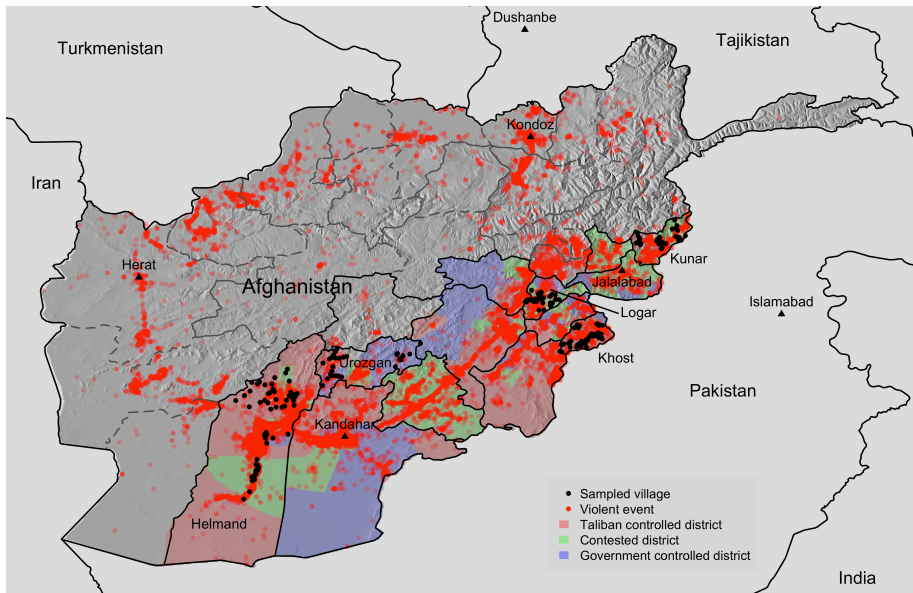
List Experiments

▶ Script for the control group:

```
I'm going to read you a list with the names of
different groups and individuals on it.  After I read
the entire list, 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
```

List Experiments

▶ Script for the treatment group:

```
I'm going to read you a list with the names of
different groups and individuals on it.  After I read
the entire list, 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; ISAF
```

The Data from the List Experiment

response value	Control Group		ISAF Treatment Group	
	frequency	proportion	frequency	proportion
0	188	20.5%	174	19.0%
1	265	28.9	278	30.3
2	265	28.9	260	28.3
3	200	21.8	182	19.8
4			24	2.6
Total	918		918	

Endorsement Experiments

▶ Script for the control group:

```
A recent proposal calls for the sweeping reform of the
Afghan prison system, including the construction of new
prisons in every district to help alleviate
overcrowding in existing facilities.  Though expensive,
new programs for inmates would also be offered, and new
judges and prosecutors would be trained.  How do you
feel about this proposal?

Strongly agree;   Agree;   Indifferent;   Disagree;
Strongly disagree;   Don't Know;   Refuse to answer
```
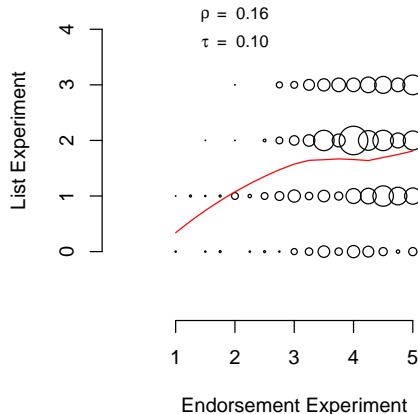
Endorsement Experiments

▶ Script for the treatment group:

A recent proposal **by ISAF** calls for the sweeping reform
of the Afghan prison system, including the construction
of new prisons in every district to help alleviate
overcrowding in existing facilities. Though expensive,
new programs for inmates would also be offered, and new
judges and prosecutors would be trained. How do you
feel about this proposal?

Strongly agree;   Agree;   Indifferent;   Disagree;
Strongly disagree;   Don't Know;   Refuse to answer

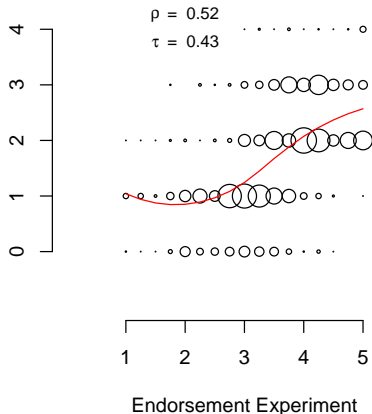# Data from the Endorsement Experiments
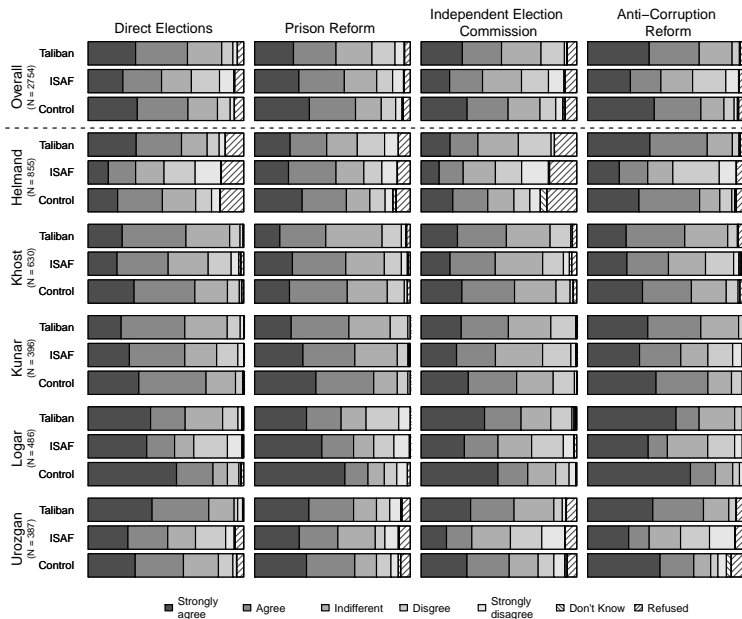
Descriptive Comparison: Overall

- ► Need for validation $\implies$ Multiple measurement strategy
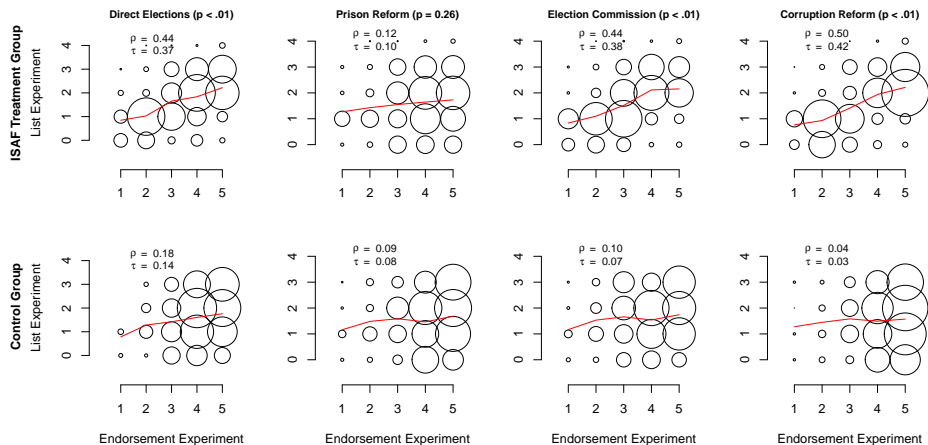- ► Two measures should give similar results

Control Group

ρ = 0.16
τ = 0.10

List Experiment (y-axis: 0, 1, 2, 3, 4)

Endorsement Experiment (x-axis: 1, 2, 3, 4, 5)

ISAF Treatment Group

ρ = 0.52
τ = 0.43

Endorsement Experiment (x-axis: 1, 2, 3, 4, 5)

Descriptive Comparison: Question by Question

Comparing and Combining List and Endorse Experiments

- Formal comparison and integration

- What is the probability of supporting ISAF?
    1. List: prob. of saying yes to the sensitive item
    2. Endorsement: prob. of endorsement having a positive effect on support for policy

- These probabilities should be similar!

- They can be estimated with a new multivariate regression method

- Endorsement and list experiments can even be combined for a joint analysis

List Experiments Framework

- N respondents
- J control items
- T_i: binary treatment indicator ($1 = $ treatment, $0 = $ control)
- V_i: pre-treatment covariates
- Y_i: outcome variable

- Define a type of each respondent by
  - total number of yes for J control items $Y_i(0)$
  - truthful answer to the sensitive item Z_i^*: $Y_i(1) = Z_i^* + Y_i(0)$
  - A total of $(2 \times (J+1))$ types

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	(0,1) (1,0)	(1,1) (1,0)
0	(0,0)	(0,1) (0,0)

List Experiments Framework

- N respondents
- J control items
- T_i: binary treatment indicator ($1 =$ treatment, $0 =$ control)
- X_i: pre-treatment covariates
- Y_i: outcome variable

- Define a type of each respondent by
    - total number of yes for J control items $Y_i(0)$
    - truthful answer to the sensitive item Z_i^*: $Y_i(1) = Z_i^* + Y_i(0)$
    - A total of $(2 \times (J+1))$ types

Y_i	Treatment group	Control group
4	(3,1)	
3	(2,1) (3,0)	(3,1) (3,0)
2	(1,1) (2,0)	(2,1) (2,0)
1	(0,1) (1,0)	(1,1) (1,0)
0	(0,0)	(0,1) (0,0)

- *Joint distribution* of $(Y_i(0), Z_i^*)$ is identified

Statistical Modeling for List Experiments

▶ Model for sensitive item: e.g., probit regression

$$\Pr(Z_i^* = 1 \mid V_i) \;=\; \Phi(V_i^\top \delta)$$

▶ Model for control items given the response to sensitive item: e.g., binomial or beta-binomial probit regression

$$\Pr(Y_i(0) = y \mid V_i, Z_i^* = z) \;=\; J \times \Phi(V_i^\top \psi_z)$$

▶ Maximum likelihood with the EM algorithm or Bayes with MCMC

Endorsement Experiments Framework

- N respondents

- J policy questions

- $Y_{ij} \in \{0, 1\}$: response of respondent i to policy j (can be ordinal)

- $T_{ij} \in \{0, 1\}$: random endorsement of policy j for respondent i

- V_i: Covariates measured prior to the treatment

Statistical Modeling for Endorsement Experiments

▶ Multiple questions $\implies$ item response theory

$$\Pr(Y_{ij} = 1 \mid T_i = t) \ = \ \Phi(\alpha_j + \beta_j(x_i + ts_{ij}^*))$$

- ▶ α_j: average popularity of policy j
- ▶ β_j: how much policy j differentiates pro- and anti-reform respondents
- ▶ x_i: "ideal point" $=$ how pro-reform respondent i is
- ▶ s_{ij}^*: endorsement effect

▶ Support level:

$$s_{ij} \ = \ \left\{ \begin{array}{ll} s_{ij}^* & \text{if } \beta_j \geq 0 \\ -s_{ij}^* & \text{otherwise} \end{array} \right.$$

such that $\frac{\partial}{\partial s_{ij}} \Pr(Y_{ij} = 1 \mid T_{ij} = 1) > 0$

Comparing and Combining the Two Models

- Key quantity: Probability of being a supporter
- List experiments:

$$\Pr(Z_i^* = 1 \mid V_i) \;\; = \;\; \Phi(V_i^\top \gamma)$$
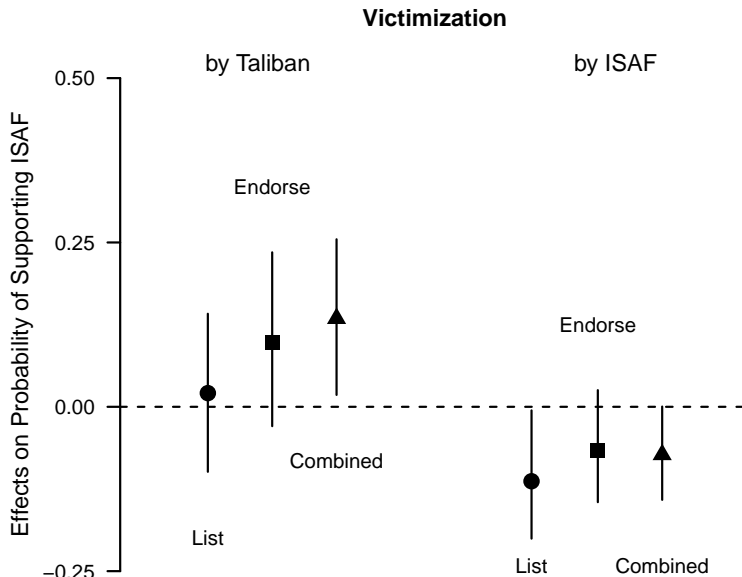
- Endorsement experiments:

$$\Pr(s_{ij} > 0 \mid V_i) \;\; = \;\; \Phi(V_i^\top \lambda / \omega)$$

- Compare the coefficients: γ and λ/ω
- Combine the two models: $\gamma = \lambda/\omega$

Overall Proportion of ISAF Supporters
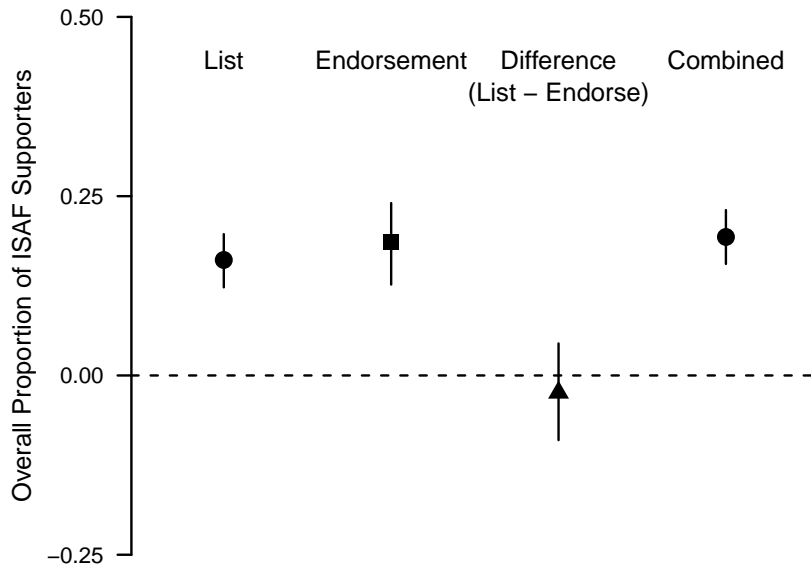
Victimization

The Mississippi Validation Study (Rosenfeld, Imai & Shapiro 2016)

- Estimate voting on anti-abortion referendum using:
  - direct question
  - list experiment (item/unmatched count technique)
  - endorsement experiment
  - randomized response

- Validate estimates against official election outcome:
  - sample from voter history file
  - county-level voting recap reports for validation

- Case selection:
  - a poll conducted 24 hours before the election predicts 44% no votes
  - the amendment was defeated 58% to 42%

- Findings:
  - direct question ⤳ significant under-estimation though efficient
  - indirect methods ⤳ much less biased though more variable
  - endorsement and randomized response ⤳ least bias

Did you vote YES or NO on the Personhood Initiative, which
appeared on the November 2011 Mississippi General Election
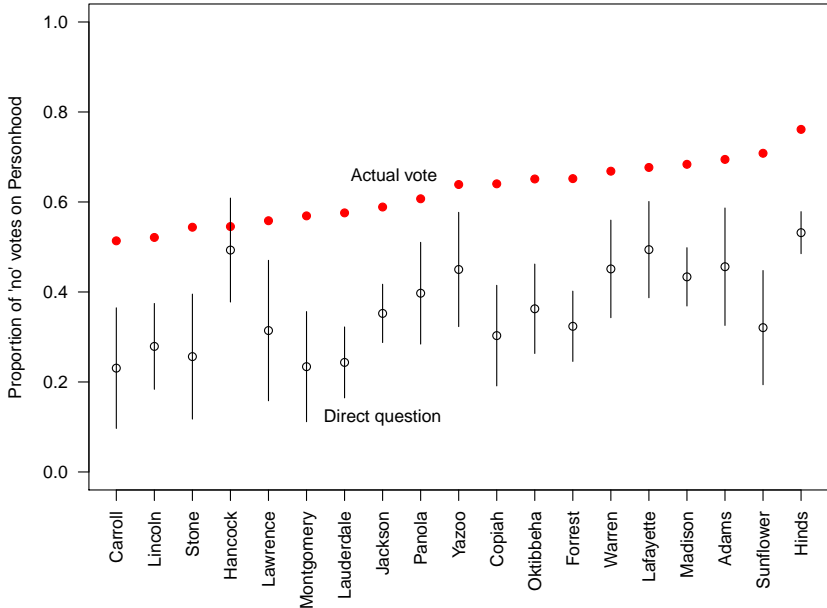ballot?

    Voted Yes

    Voted No

    Did not vote

    Don't know

    Refused

Bias of the Direct Question

List Experiment

Here is a list of four things that some people have done and some people have not. Please listen to them and then tell me HOW MANY of them you have done in the past two years. Do not tell me which you have and have not done. Just tell me how many:

Discussed politics with family or friends

Cast a ballot for Governor Phil Bryant

Paid dues to a union

Given money to a Tea Party candidate or organization

(treatment) Voted 'YES' on the 'Personhood' Initiative

How many of these things have you done in the past two years?

Endorsement Experiment

We'd like to get your overall opinion of some people in the news. As I read each name, please say if you have a very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable opinion of each person.

(control) Phil Bryant, Governor of Mississippi?

(treatment) Phil Bryant, Governor of Mississippi, who campaigned in favor of the 'Personhood' Initiative on the 2011 Mississippi General Election ballot?

Randomized Response

To answer this question, you will need a coin. Once you have
found one, please toss the coin two times and note the results of
those tosses (heads or tails) one after the other on a sheet of
paper. Do not reveal to me whether your coin lands on heads or
tails. After you have recorded the results of your two coin
tosses, just tell me you are ready and we will begin.

Now, please answer 'yes' if either your second coin toss came up
heads or you voted 'YES' on the Personhood Initiative, which
appeared on the November 2011 Mississippi General Election
ballot.

Yes

No

Don't know

Refused
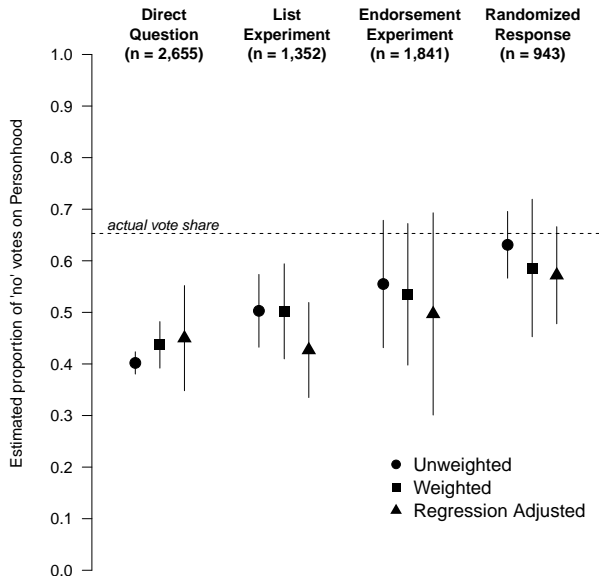
County-level Analysis
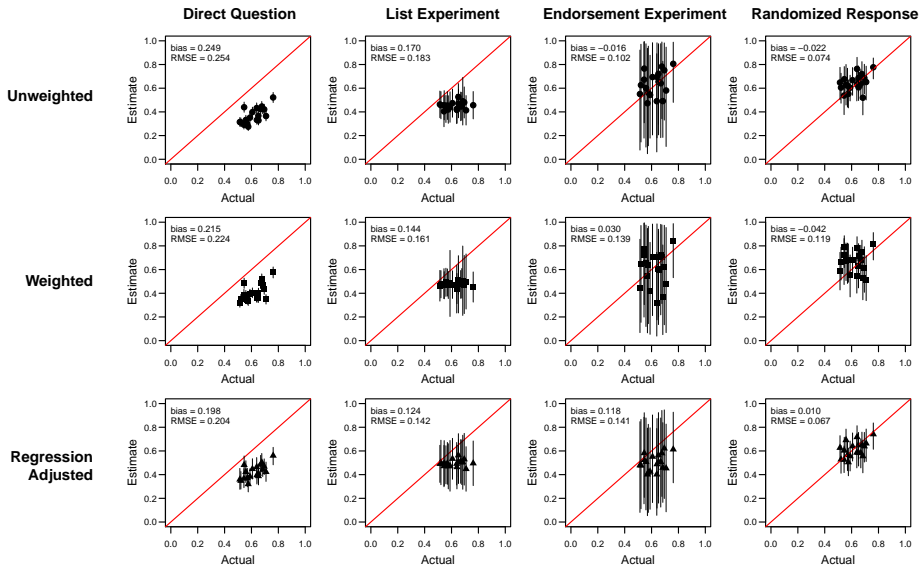
The Proposed Methodology for List Experiment

▶ List experiment can be analyzed by method of moments:

$$\mathbb{E}(Y_i \mid T_i, X_i) \;=\; \underbrace{f(X_i, \gamma)}_{\text{Control Items}} + T_i \; \underbrace{g(X_i, \delta)}_{\text{Sensitive Trait}}$$

▶ We simply add moment conditions of the form

$$\mathbb{E}[g(X_i, \delta)] \;=\; h$$
$$\mathbb{E}[g(X_i, \delta) \mid G_i = k] \;=\; h_k$$

▶ A similar strategy works for randomized experiment

▶ (Testable) Assumption: Same parameters solve all moment conditions
 ⤳ Constant parameters across groups.

Proposed Methodology for Endorsement Experiment

1. Use precinct-level indicators as covariates

$$s_{ij}^* \overset{\text{indep.}}{\sim} \mathcal{N}(\lambda^\top X_i, \omega^2),$$

2. Assume the prior for precinct random effects λ_r:

$$\lambda_r \overset{\text{indep.}}{\sim} \mathcal{N}(\mu_{\lambda_{\text{county}[r]}}, \sigma^2_{\text{county}[r]}),$$

where county$[r]$ denotes the county containing precinct r.

3. Choose $\mu_{\lambda_{\text{county}[r]}}$ to match auxiliary information

   - We assume the following conjugate prior for ω^2:
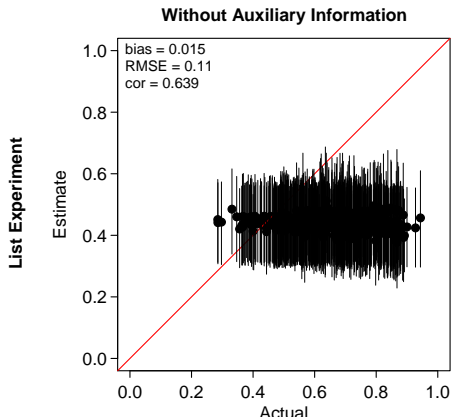
   $$\omega^2 \sim \kappa/\chi^2_\nu$$

   - The marginal prior for s_{ij}^* is a t-distribution:

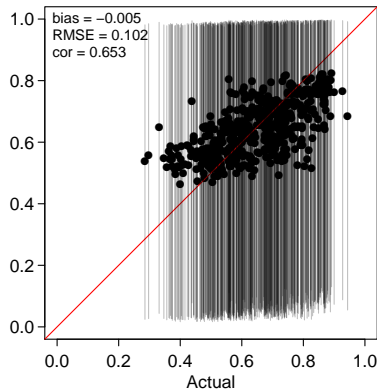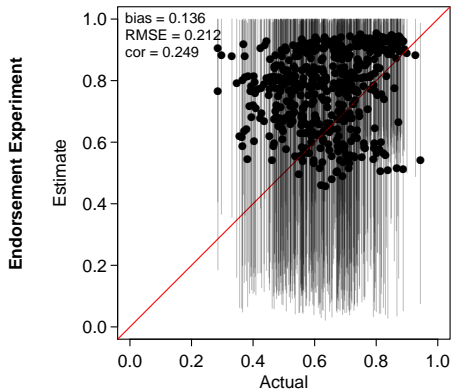 $$s_{ij}^* \mid G_i = k \overset{\text{indep.}}{\sim} t_\nu(\mu_{\lambda_k}, \sigma^2_k).$$

Efficiency Comparison with Direct Questioning

	List Experiment N=1,325		Randomized Response N=818		Endorsement Experiment N=1,841	
	s.e.	ratio	s.e.	ratio	s.e.	ratio
Direct questioning	0.017		0.021		0.289	
No auxiliary info.	0.067	3.963	0.040	1.928	0.348	1.204
With auxiliary info.	0.019	1.150	0.018	0.855	0.276	0.955

Auxiliary Information Improves List Experiment

Without Auxiliary Information

bias = 0.015
RMSE = 0.11
cor = 0.639

With Auxiliary Information

bias = 0.042
RMSE = 0.098
cor = 0.774

List Experiment / Estimate — Actual
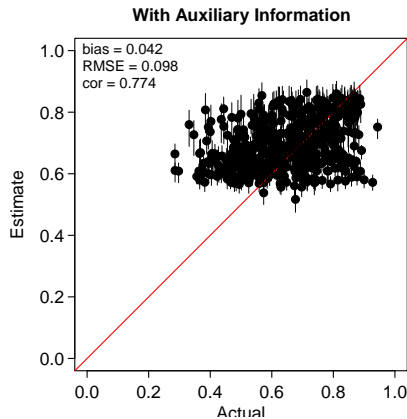
Auxiliary Information Improves Endorsement Experiment
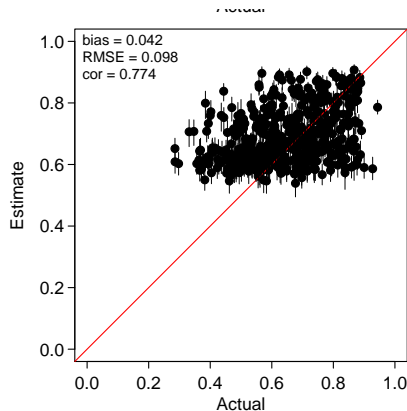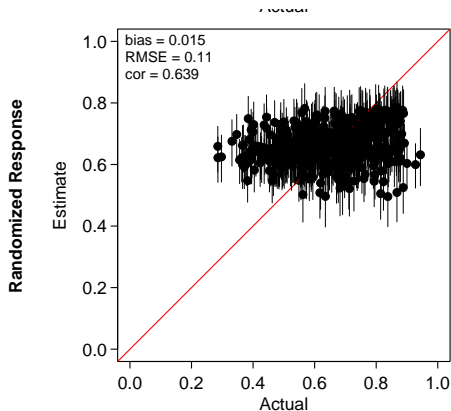
Auxiliary Information Improves Multivariate Inference
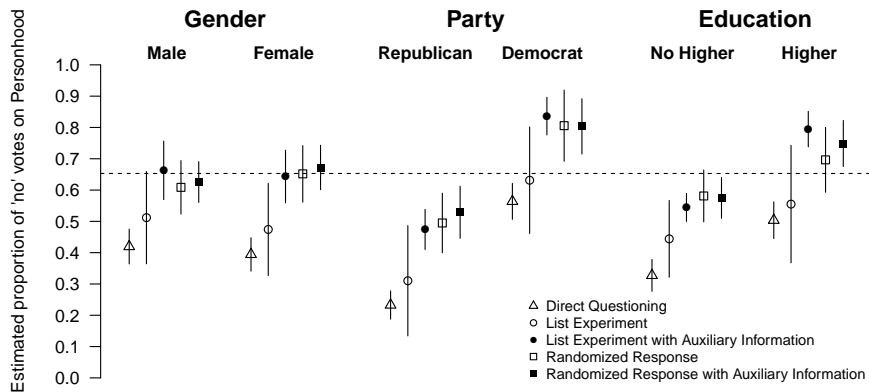
Concluding Remarks

- Direct question is severely biased

- All indirect methods reduce bias:
    - Endorsement and randomized response ⤳ least bias
    - List experiment ⤳ ceiling/floor effects, design effects
    - Ease of implementation: list > endorse > randomized response
- But, they are inefficient: bias-variance tradeoff

- Two ways to improve indirect question methods:
    1. Use of multiple-measurement strategies when truth is not available
    2. Use aggregate-level truth to improve individual-level estimates

- Open-source software:
    - `list` for list experiment (Blair, Imai & Park)
    - `endorse` for endorsement experiment (Shiraito & Imai)
    - `rr` for randomized response (Blair, Imai & Zhou)