# Eliciting Truthful Answers to Sensitive Survey Questions: New Statistical Methods for List and Endorsement Experiments

Kosuke Imai

Princeton University

September 15, 2010

# Project Reference

- PAPERS:
  1. Imai. "Statistical Inference for the Item Count Technique." Working Paper.
  2. Blair and Imai. "Statistical Analysis of List Experiments." Work in progress.
  3. Bullock, Imai, and Shapiro. "Measuring Political Support and Issue Ownership Using Endorsement Experiments, with Application to the Militant Groups in Pakistan." Working Paper.
- CODE AND SOFTWARE:
  1. Code for endorsement experiments available at the Dataverse
  2. Blair, Graeme, and Kosuke Imai. `list: Multivariate Statistical Analysis for the Item Count Technique.` an R package
- PROJECT WEBSITE:

http://imai.princeton.edu/projects/sensitive.html

# Motivation

- Survey is used widely in social sciences
- Validity of survey depends on the accuracy of self-reports
- Sensitive questions $\Longrightarrow$ social desirability, privacy concerns
  e.g., racial prejudice, corruptions
- Lies and nonresponses

- How can we elicit truthful answers to sensitive questions?
- Survey methodology: protect privacy through indirect questioning
- Statistical methodology: efficiently recover underlying responses

# Survey Techniques for Sensitive Questions

- Randomized Response Technique (Warner, 1965)
  - Most extensively studied and commonly used
  - Use randomization to protect privacy
  - Difficulties: logistics, lack of understanding among respondents
- **Item Count Technique** (Miller, 1984)
  - Also known as list experiment and unmatched count technique
  - Use aggregation to protect privacy
  - Develop new estimators to enable *multivariate regression analysis*
  - Application: racial prejudice in the US
- **Endorsement Experiments**
  - Use randomized endorsements to measure support levels
  - Develop a measurement model based on *item response theory*
  - Application: Pakistanis' support for Islamic militant groups

# Item Count Technique: Example

- The 1991 National Race and Politics Survey
- Randomize the sample into the treatment and control groups
- The script for the control group

```
Now I'm going to read you three things that sometimes
make people angry or upset. After I read all three,
just tell me HOW MANY of them upset you. (I don't
want to know which ones, just how many.)

(1) the federal government increasing the tax on
gasoline;
(2) professional athletes getting million-dollar-plus
salaries;
(3) large corporations polluting the environment.
```

# Item Count Technique: Example

- The 1991 National Race and Politics Survey
- Randomize the sample into the treatment and control groups
- The script for the treatment group

```
Now I'm going to read you four things that sometimes
make people angry or upset. After I read all four,
just tell me HOW MANY of them upset you. (I don't
want to know which ones, just how many.)

(1) the federal government increasing the tax on
gasoline;
(2) professional athletes getting million-dollar-plus
salaries;
(3) large corporations polluting the environment;
(4) a black family moving next door to you.
```

# Design Considerations and Standard Analysis

- Privacy is protected unless respondents' truthful answers are yes for all sensitive and non-sensitive items $\Longrightarrow$ underestimation
- A large number of non-sensitive items yields high variance
- Less efficient than direct questioning
- Negative correlation across non-sensitive items is desirable

- Standard difference-in-means estimator:

$$\hat{\tau} = \textit{treatment group mean} - \textit{control group mean}$$

- Unbiased for the population proportion
- Stratification is possible on discrete covariates
  - large sample size is required
  - not desirable for continuous covariates
- No existing method allows for multivariate regression analysis

# Two-Step Nonlinear Least Squares (NLS) Estimator

- Generalize the difference-in-means estimator to a multivariate regression estimator
- The Model:

$$Y_i \;=\; f(X_i, \gamma) + T_i g(X_i, \delta) + \epsilon_i$$

  - $Y_i$: response variable
  - $T_i$: treatment variable
  - $X_i$: covariates
  - $f(x, \gamma)$: model for non-sensitive items, e.g., $J \times \text{logit}^{-1}(x^\top \gamma)$
  - $g(x, \delta)$: model for sensitive item, e.g., $\text{logit}^{-1}(x^\top \delta)$
- Two-step estimation procedure:
  1. Fit the $f(x, \gamma)$ model to the control group via NLS and obtain $\hat{\gamma}$
  2. Fit the $g(x, \delta)$ model to the treatment group via NLS after subtracting $f(X_i, \hat{\gamma})$ from $Y_i$ and obtain $\hat{\delta}$
- Standard errors via the method of moments
- When no covariate, it reduces to the difference-in-means estimator

# Extracting More Information from the Data

- Define a "type" of each respondent by $(Y_i(0), Z_{i,J+1})$
  - $Y_i(0)$: total number of yes for non-sensitive items $\in \{0, 1, \ldots, J\}$
  - $Z_{i,J+1}$: truthful answer to the sensitive item $\in \{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 2$)

| $Y_i$ | Treatment group | Control group |
|---|---|---|
| 3 | (2,1) | |
| 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* is identified

# Extracting More Information from the Data

- Define a "type" of each respondent by $(Y_i(0), Z_{i,J+1})$
  - $Y_i(0)$: total number of yes for non-sensitive items $\{0, 1, \ldots, J\}$
  - $Z_{i,J+1}$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 2$)

| $Y_i$ | Treatment group | Control group |
|---|---|---|
| 3 | (2,1) | |
| 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* is identified:

$$\Pr(\text{type} = (y, 1)) = \Pr(Y_i \leq y \mid T_i = 0) - \Pr(Y_i \leq y \mid T_i = 1)$$

# Extracting More Information from the Data

- Define a "type" of each respondent by $(Y_i(0), Z_{i,J+1})$
  - $Y_i(0)$: total number of yes for non-sensitive items $\{0, 1, \ldots, J\}$
  - $Z_{i,J+1}$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 2$)

| $Y_i$ | Treatment group | Control group |
|---|---|---|
| 3 | (2,1) | |
| 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* is identified:

$$\Pr(\text{type} = (y, 1)) = \Pr(Y_i \le y \mid T_i = 0) - \Pr(Y_i \le y \mid T_i = 1)$$

$$\Pr(\text{type} = (y, 0)) = \Pr(Y_i \le y \mid T_i = 1) - \Pr(Y_i < y \mid T_i = 0)$$

# The Likelihood Function

- $g(x, \delta)$: model for sensitive item, e.g., logistic regression
- $h_z(y; x, \psi_z) = \Pr(Y_i(0) = y \mid X_i = x, Z_{i,J+1} = z)$ :
  model for non-sensitive item given the response to sensitive item, e.g., binomial or beta-binomial regression
- Likelihood:

$$\prod_{i \in \mathcal{J}(1,0)} (1 - g(X_i, \delta)) h_0(0; X_i, \psi_0) \prod_{i \in \mathcal{J}(1,J+1)} g(X_i, \delta) h_1(J; X_i, \psi_1)$$

$$\times \prod_{y=1}^{J} \prod_{i \in \mathcal{J}(1,y)} \{g(X_i, \delta) h_1(y-1; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

$$\times \prod_{y=0}^{J} \prod_{i \in \mathcal{J}(0,y)} \{g(X_i, \delta) h_1(y; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

where $\mathcal{J}(t, y)$ represents a set of respondents with $(T_i, Y_i) = (t, y)$

- It would be a *nightmare* to maximize this!

# Missing Data Framework and the EM Algorithm

- Consider $Z_{i,J+1}$ as missing data
- For some respondents, $Z_{i,J+1}$ is known
- The complete-data likelihood has a much *simpler* form:

$$\prod_{i=1}^{N} \left\{ g(X_i, \delta) h_1(Y_i - 1; X_i, \psi_1)^{T_i} h_1(Y_i; X_i, \psi_1)^{1-T_i} \right\}^{Z_{i,J+1}} \times \{(1 - g(X_i, \delta)) h_0(Y_i; X_i, \psi_0)\}^{1-Z_{i,J+1}}$$

- The EM algorithm: only separate optimization of $g(x, \delta)$ and $h_z(y; x, \psi_z)$ is required
  - weighted logistic regression
  - weighted binomial or beta-binomial regression
- Asymptotically unbiased and most efficient

# Empirical Application: Racial Prejudice in the US

- Kuklinski *et al*. (1997) analyzes the 1991 National Race and Politics survey with the standard difference-in-means estimator
- Finding: Southern whites are more prejudiced against blacks than non-southern whites – no evidence for the "New South"
- The limitation of the original analysis:
  *"So far our discussion has implicitly assumed that the higher level of prejudice among white southerners results from something uniquely "southern," what many would call southern culture. This assumption could be wrong. If white southerners were older, less educated, and the like – characteristics normally associated with greater prejudice – then demographics would explain the regional difference in racial attitudes, leaving culture as little more than a small and relatively insignificant residual."*
- Need for a multivariate regression analysis

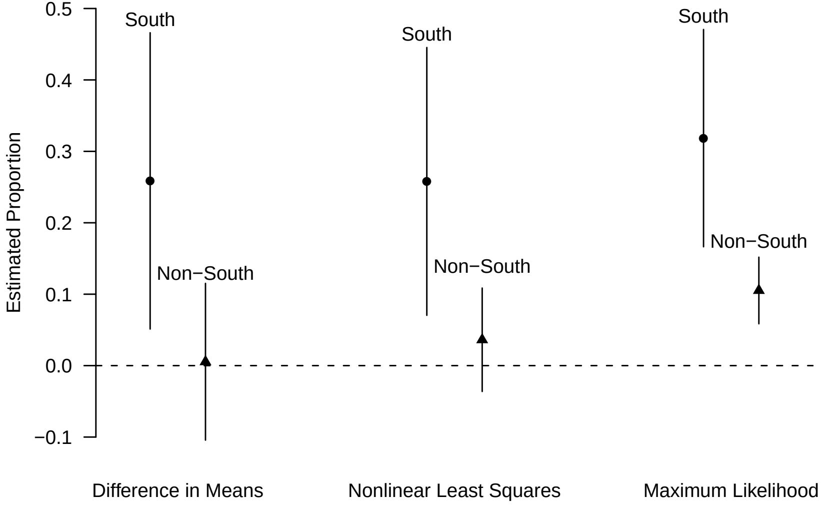
# Results of the Multivariate Analysis

- Logistic regression model for sensitive item
- Binomial regression model for non-sensitive item (not shown)
- Little over-dispersion
- Likelihood ratio test supports the constrained model

| | **Nonlinear Least Squares** | | **Maximum Likelihood** | | | |
|---|---|---|---|---|---|---|
| | | | Constrained | | Unconstrained | |
| Variables | est. | s.e. | est. | s.e. | est. | s.e. |
| Intercept | −7.084 | 3.669 | −5.508 | 1.021 | −6.226 | 1.045 |
| South | 2.490 | 1.268 | 1.675 | 0.559 | 1.379 | 0.820 |
| Age | 0.026 | 0.031 | 0.064 | 0.016 | 0.065 | 0.021 |
| Male | 3.096 | 2.828 | 0.846 | 0.494 | 1.366 | 0.612 |
| College | 0.612 | 1.029 | −0.315 | 0.474 | −0.182 | 0.569 |

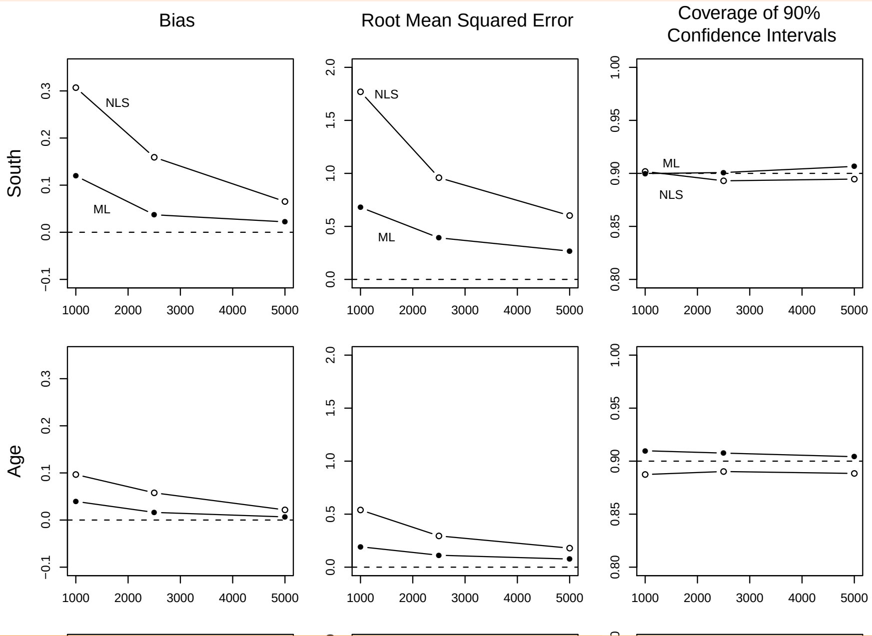
- The original conclusion is supported
- Standard errors are much smaller for ML estimator

# Estimated Proportion of Prejudiced Whites




- Regression adjustments and MLE yield more efficient estimates

# Simulation Evidence



# When Can List Experiments Fail?

- Two possibilities:
  1. Respondents give different answers to non-sensitive items depending on whether the list includes a sensitive item
  2. Respondents lie about the sensitive item
     - Ceiling effects: too many yeses for non-sensitive items
     - Floor effects: too many noes for non-sensitive items
- Both types of failure are difficult to detect from observed data
- Importance of pilot studies: ask non-sensitive and sensitive items separately and compare the resopnses
- Question: Can these failures be addressed statistically?

# Hypothesis Tests for Detecting a Failure

- Recall the table of respondent types
- Under our assumption, we have a non-negative proportion for each type

$$\Pr(\text{type} = (y, 1)) = \Pr(Y_i \le y \mid T_i = 0) - \Pr(Y_i \le y \mid T_i = 1)$$

$$\Pr(\text{type} = (y, 0)) = \Pr(Y_i \le y \mid T_i = 1) - \Pr(Y_i < y \mid T_i = 0)$$

- Null hypothesis: ALL of these proportions are non-negative
- Alternative hypothesis: At least one is negative
- Proposed statistical tests:
  - The Bonferroni test: finite sample, nonparametric
  - The union-intersection test: asymptotic approximation
  - The Wald test: asymptotic approximation
- Failure to reject the null may arise from the lack of power

# Modeling Ceiling and Floor Effects

- Potential liars:

| $Y_i$ | Treatment group | Control group |
|---|---|---|
| 4 | (3,1) | |
| 3 | (2,1) (3,0) (3,1)* | (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)* | (0,1) (0,0) |

- Previous tests do not detect these liars: proportions would still be positive so long as they are truthful about non-sensitive items
- Proposed strategy: model ceiling and/or floor effects under an additional assumption
- Aditional assumption: conditional independence between items given covariates
- ML estimation can be extended to this situation

# Endorsement Experiments

- Measuring support for political actors (e.g., candidates, parties) when studying sensitive questions
- Ask respondents to rate their support for a set of policies endorsed by randomly assigned political actors
- Experimental design:
  1. Select policy questions
  2. Randomly divide sample into control and treatment groups
  3. Across respondents and questions, randomly assign political actors for endorsement (no endorsement for the control group)
  4. Compare support level for each policy endorsed by different actors

# The Pakistani Survey Experiment

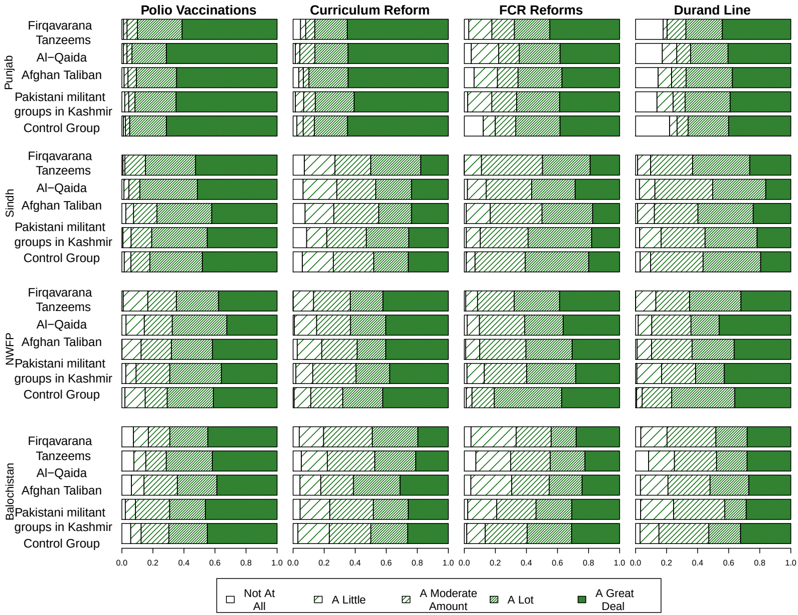
- 6,000 person urban-rural sample
- Four very different groups:
  - Pakistani militants fighting in Kashmir (a.k.a. Kashmiri tanzeem)
  - Militants fighting in Afghanistan (a.k.a. Afghan Taliban)
  - Al-Qa'ida
  - Firqavarana Tanzeems (a.k.a. sectarian militias)
- Four policies:
  - WHO plan to provide universal polio vaccination across Pakistan
  - Curriculum reform for religious schools
  - Reform of FCR to make Tribal areas equal to rest of the country
  - Peace jirgas to resolve disputes over Afghan border (Durand Line)
- Response rate over 90%

# Endorsement Experiment Questions: Example

- The script for the control group
- `The World Health Organization recently announced a plan to introduce universal Polio vaccination across Pakistan. How much do you support such a plan?`

- The script for the treatment group
- `The World Health Organization recently announced a plan to introduce universal Polio vaccination across Pakistan, a policy that has received support from Al-Qa'ida. How much do you support such a plan?`

# Distribution of Responses



# Endorsement Experiments Framework

- Data from an endorsement experiment:
  - $N$ respondents
  - $J$ policy questions
  - $K$ political actors
  - $Y_{ij} \in \{0, 1\}$: response of respondent $i$ to policy question $j$
  - $T_{ij} \in \{0, 1, \ldots, K\}$: political actor randomly assigned to endorse policy $j$ for respondent $i$
  - $Y_{ij}(t)$: potential response given the endorsement by actor $t$
  - Covariates measured prior to the treatment

- Quadratic random utility model:

$$U_i(\zeta_{j1}, k) = -\|(x_i + s_{ijk}) - \zeta_{j1}\|^2 + \eta_{ij},$$
$$U_i(\zeta_{j0}, k) = -\|(x_i + s_{ijk}) - \zeta_{j0}\|^2 + \nu_{ij},$$

where $x_i$ is the ideal point and $s_{ijk}$ is the support level

- The statistical model (item response theory):

$$\Pr(Y_{ij} = 1 \mid T_{ij} = k) = \Pr(Y_{ij}(k) = 1) = \Pr(U_i(\zeta_{j1}, k) > U_i(\zeta_{j0}, k))$$
$$= \Pr(\alpha_j + \beta_j(x_i + s_{ijk}) > \epsilon_{ij})$$

- Hierarchical modeling:

$$x_i \overset{\text{indep.}}{\sim} \mathcal{N}(Z_i^\top \delta,\ \sigma_x^2)$$
$$s_{ijk} \overset{\text{indep.}}{\sim} \mathcal{N}(Z_i^\top \lambda_{jk},\ \omega_{jk}^2)$$
$$\lambda_{jk} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta_k, \Phi_k)$$

- "Noninformative" hyper prior on $(\alpha_j, \beta_j, \delta, \theta_k, \omega_{jk}^2, \Phi_k)$

# Quantities of Interest and Model Fitting

- Average support level for each militant group $k$

$$\begin{aligned} \tau_{jk}(Z_i) &= Z_i^\top \lambda_{jk} && \text{for each policy } j \\ \kappa_k(Z_i) &= Z_i^\top \theta_k && \text{averaging over all policies} \end{aligned}$$

- Standardize them by dividing the (posterior) standard deviation of ideal points

- Bayesian Markov chain Monte Carlo algorithm
- Multiple chains to monitor convergence
- Implementation via `JAGS` (Plummer)

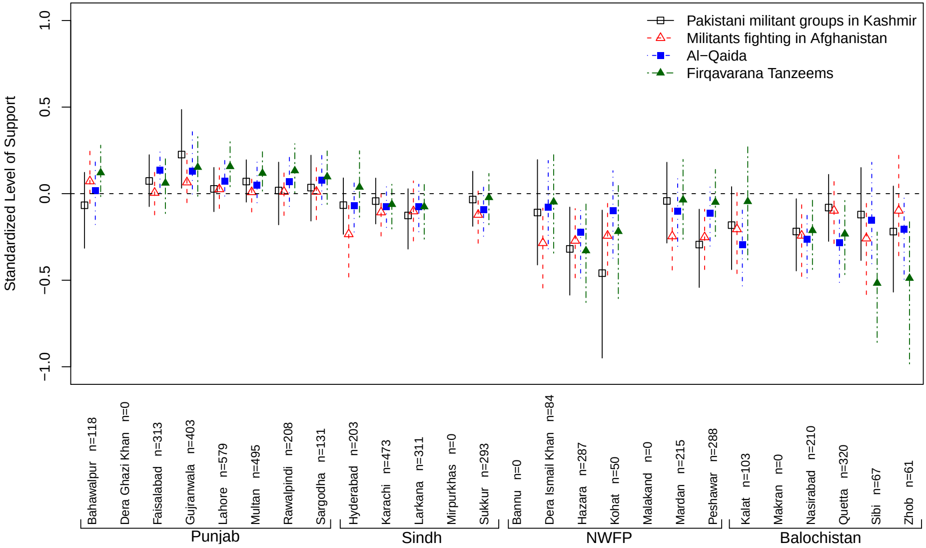
# Model for the Division Level Support

- Ordered response with an intercept $\alpha_{jl}$ varying across divisions
- The model specification:

$$
\begin{aligned}
x_i &\overset{\text{indep.}}{\sim} \mathcal{N}(\delta_{\text{division}[i]}, 1) \\
s_{ijk} &\overset{\text{indep.}}{\sim} \mathcal{N}(\lambda_{k,\text{division}[i]}, \omega_k^2) \\
\delta_{\text{division}[i]} &\overset{\text{indep.}}{\sim} \mathcal{N}(\mu_{\text{province}[i]}, \sigma^2_{\text{province}[i]}) \\
\lambda_{k,\text{division}[i]} &\overset{\text{indep.}}{\sim} \mathcal{N}(\theta_{k,\text{province}[i]}, \Phi_{k,\text{province}[i]})
\end{aligned}
$$

- Averaging over policies
- Partial pooling across divisions within each province

# Estimated Division Level Support



# Model with Individual Covariates

- Ordered response with an intercept $\alpha_{jl}$ varying across divisions
- The model specification:

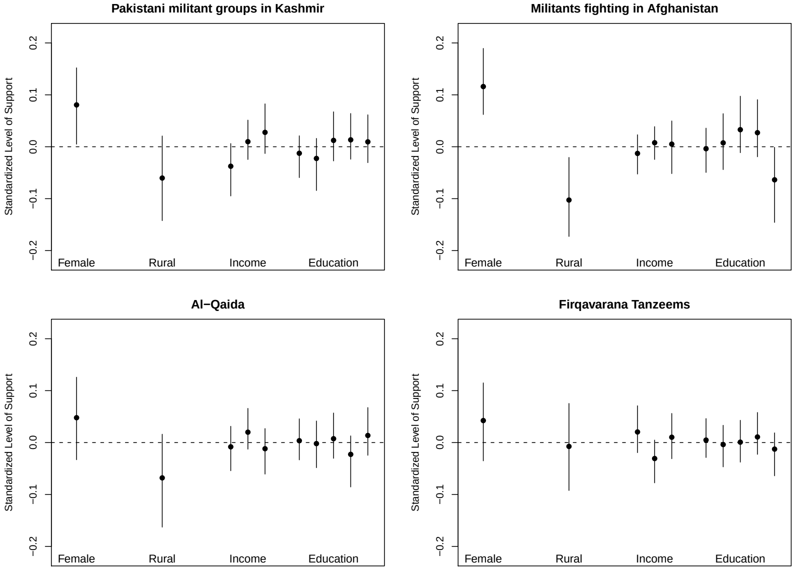
$$
\begin{aligned}
x_i &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\delta_{\text{division}[i]} + Z_i^\top \delta^Z, 1) \\
s_{ijk} &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\lambda_{k,\text{division}[i]} + Z_i^\top \lambda_k^Z, \omega_k^2) \\
\delta_{\text{division}[i]} &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\mu_{\text{province}[i]}, \sigma^2_{\text{province}[i]}) \\
\lambda_{k,\text{division}[i]} &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\theta_{k,\text{province}[i]}, \Phi_{k,\text{province}[i]})
\end{aligned}
$$

- Expands upon the division level model to include individual level covariates:

gender, urban/rural, education, income

- Individual level covariate effects after accounting for differences across divisions
- Poststratification on these covariates using the census

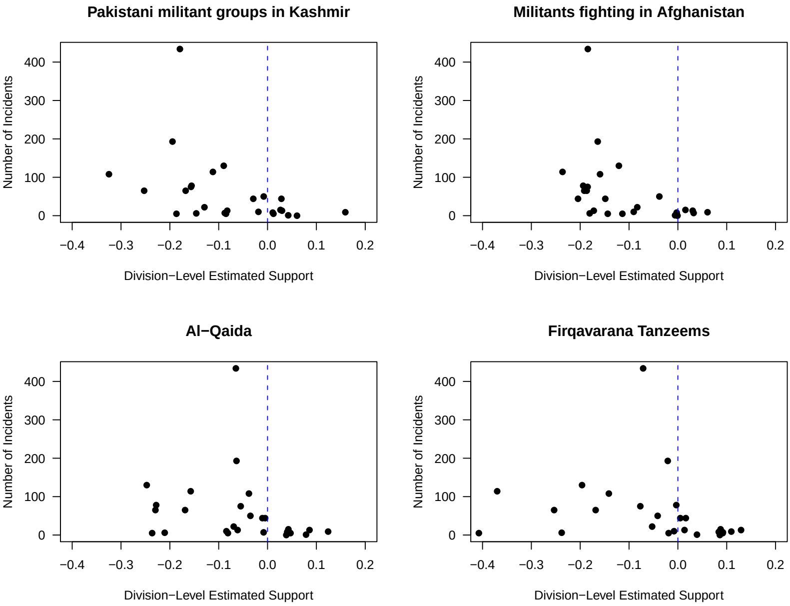
# Estimated Effects of Individual Covariates




- Demographics play a small role in explaining support for groups

# Regional Clustering of the Support for Al-Qaida

# Correlation between Support and Violence

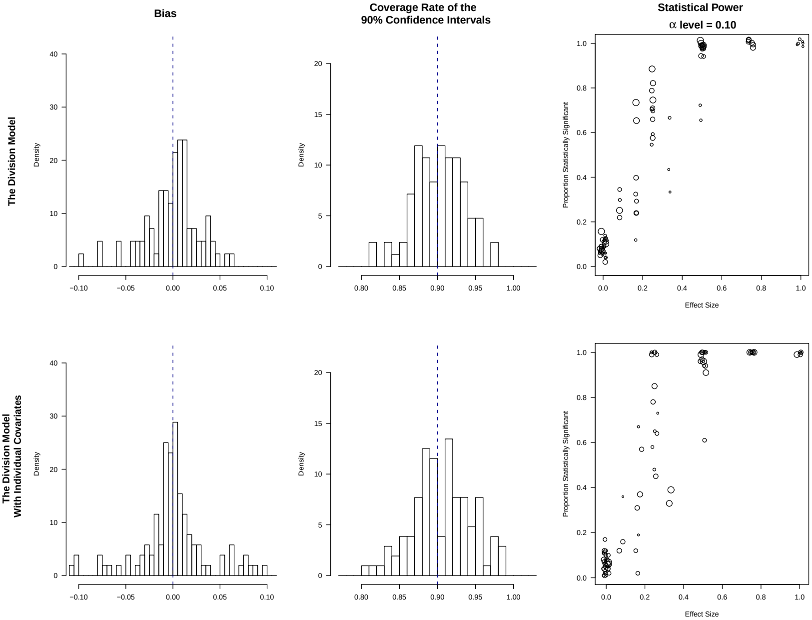


# Simulation Studies

1. Based on the Pakistani Data
   - Same 2 models plus province-level issue ownership model
   - Top-level parameters held constant across simulations
   - Sample sizes and distribution same as before
   - Ideal points, endorsements and responses follow IRT models
2. Varying sample sizes
   - Model for division-level estimates with no covariates
   - Model for province-level estimates with no covariates but support varying across policies
   - $N = 1000, 1500, 2000$
   - Again, top-level parameters held constant across simulations while ideal points, endorsements and responses follow IRT models

- 100 simulations under each scenario (3 chains, 60000 iterations)
- Frequentist evaluation of Bayesian estimators

# Monte Carlo Evidence based on the Pakistani Data



# Model for the Province Level Issue Ownership

- The Model specification:

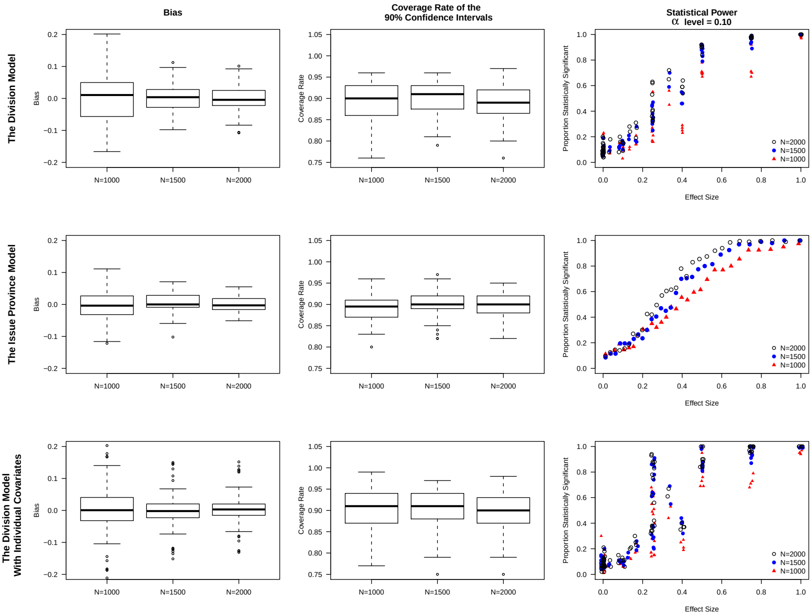
$$x_i \overset{\text{indep.}}{\sim} \mathcal{N}(\delta_{\text{province}[i]}, 1)$$

$$s_{ijk} \overset{\text{indep.}}{\sim} \mathcal{N}(\lambda_{jk,\text{province}[i]}, \omega_{jk}^2)$$

$$\lambda_{jk,\text{province}[i]} \overset{\text{indep.}}{\sim} \mathcal{N}(\theta_{k,\text{province}[i]}, \Phi_{k,\text{province}[i]})$$

- Pooling across divisions within each province
- Partial pooling across policies

# Monte Carlo Evidence with Varying Sample Size



# Concluding Remarks

- Viable alternatives to the randomized response technique
- Item Count Technique
  - Easy for researchers to implement
  - Easy for respondents to understand
  - Widely applicable
  - Need to carefully choose non-sensitive items
  - Aggregation $\Longrightarrow$ loss of efficiency
- Endorsement Experiments
  - Most indirect questioning
  - Applicability limited to measuring support
  - Need to carefully choose policy questions
  - Many groups $\Longrightarrow$ loss of efficiency
- New statistical methods for efficient inference
- Open-source software and code available