Measuring Political Support and Issue Ownership Using Endorsement Experiments, with Application to the Militant Groups in Pakistan

Kosuke Imai

Princeton University

Joint work with Will Bullock and Jacob Shapiro

November 5, 2010

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
  - Most extensively studied and commonly used
  - Use randomization to protect privacy
  - Difficulties: logistics, lack of understanding among respondents
- List Experiments
  - Also known as block total response and item 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

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

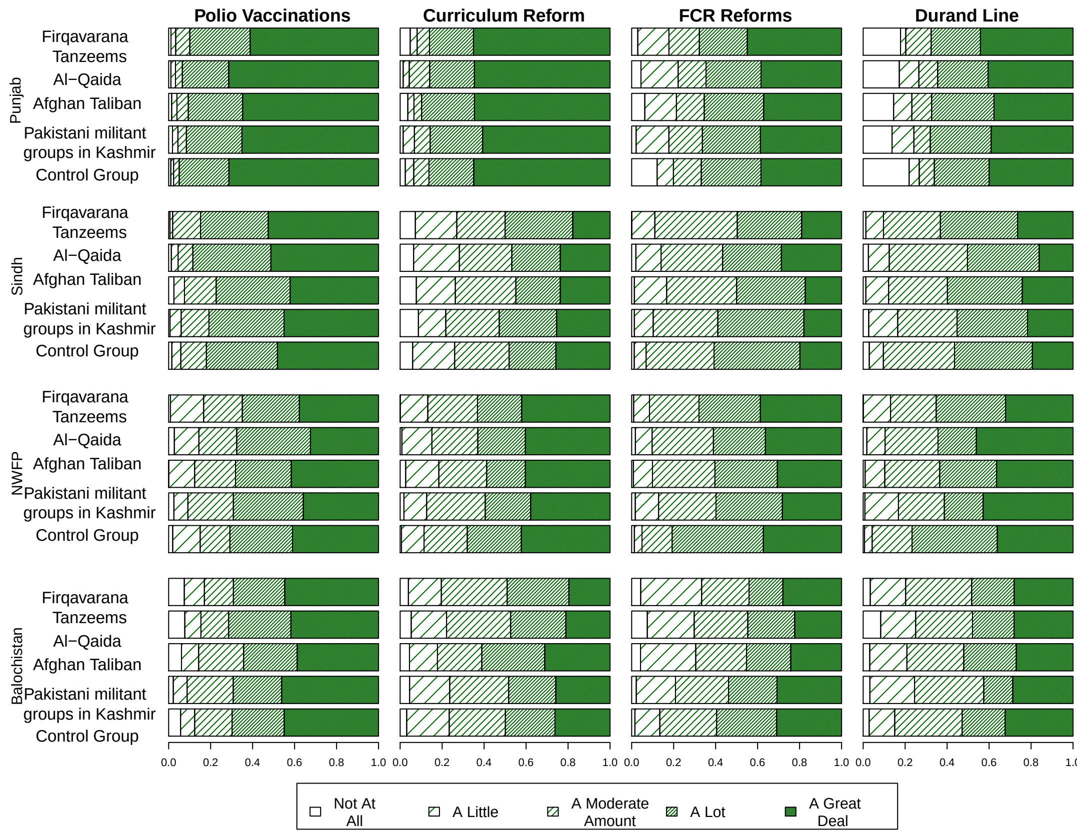
The Pakistani Survey Experiment

- 6,000 person urban-rural sample
- Four 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

The Proposed Model

- 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):

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

- Hierarchical modeling:

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

- "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

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

- 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)

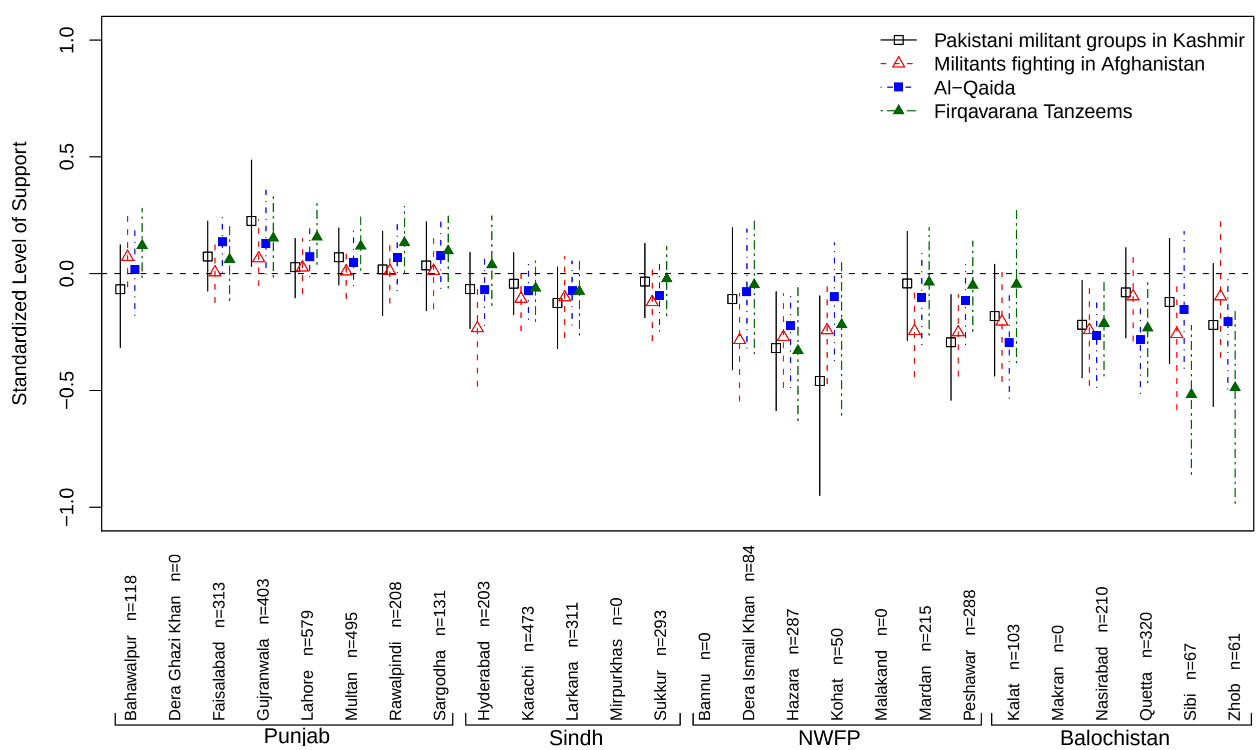
Model for the Division Level Support

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

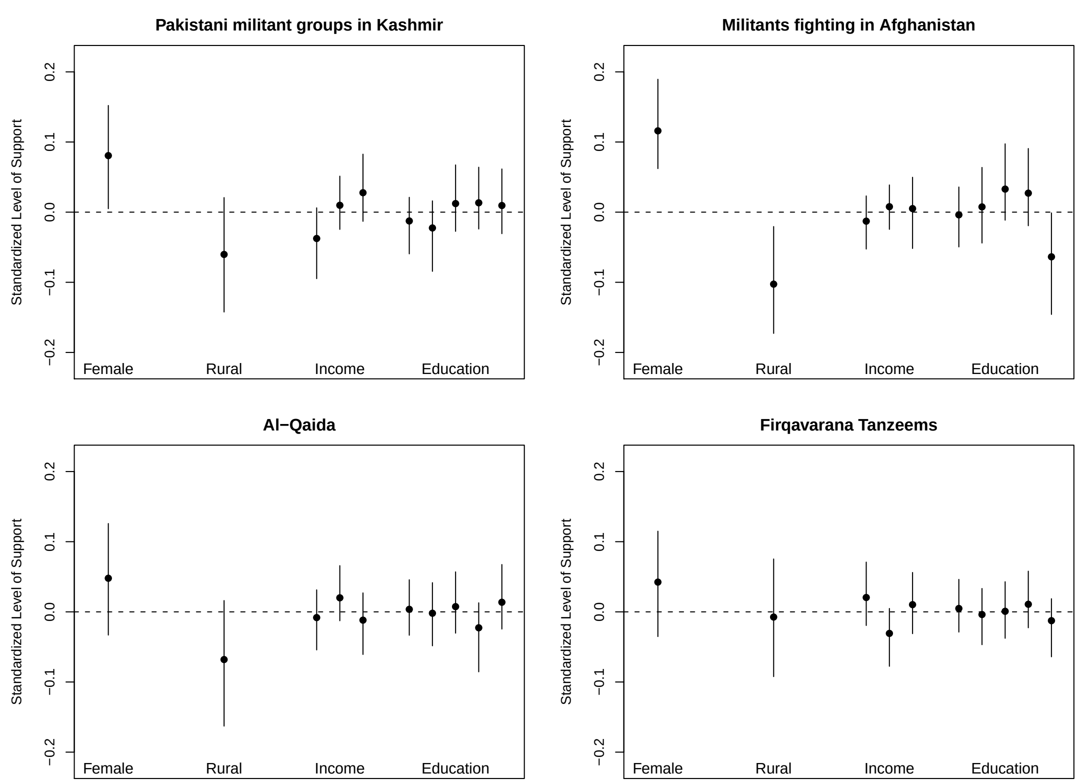
$$
\begin{aligned}
x_i &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\delta_{\text{division}[i]} + Z_i^\top \delta^Z, 1) \\
s_{ijk} &\stackrel{\text{indep.}}{\sim} \mathcal{N}(\lambda_{k,\text{division}[i]} + Z_i^\top \lambda_k^Z, \omega_k^2) \\
\delta_{\text{division}[i]} &\stackrel{\text{indep.}}{\sim}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

- 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

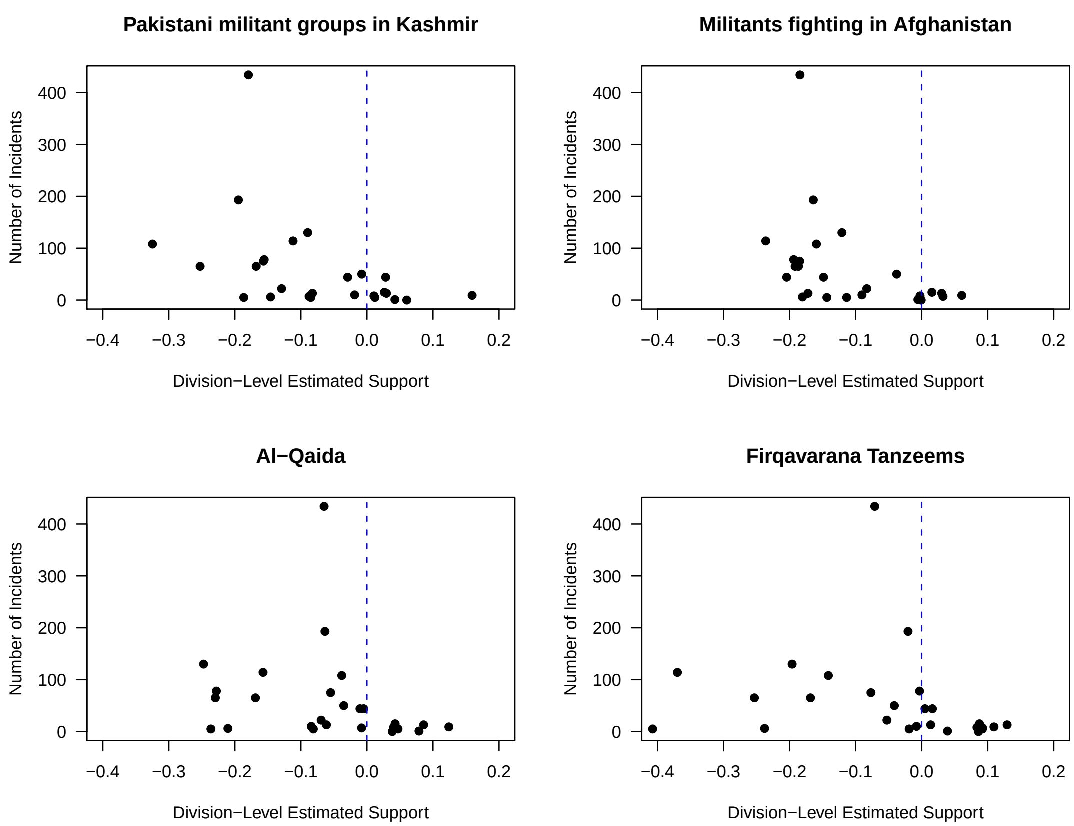
Estimated Effects of Individual Covariates



- Demographics play a small role in explaining support for groups

Regional Clustering of the Support for Al-Qaida

Correlation between Support and Violence

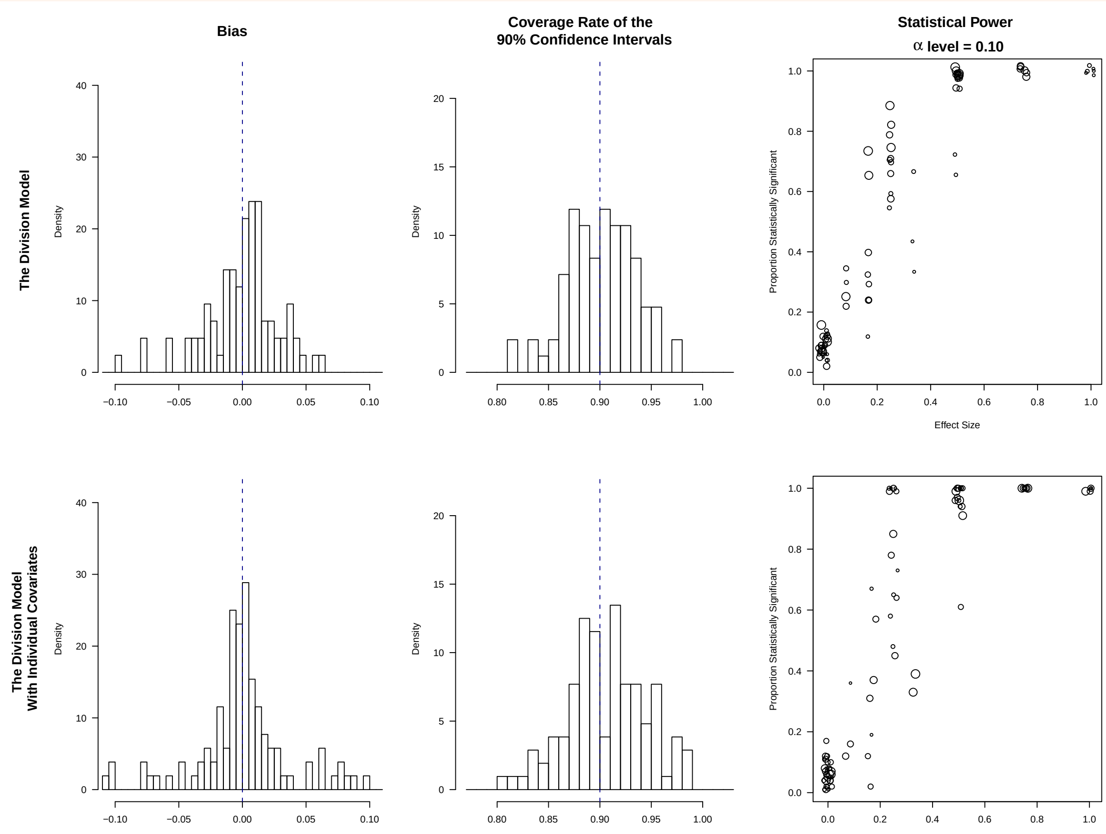


Simulation Studies

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



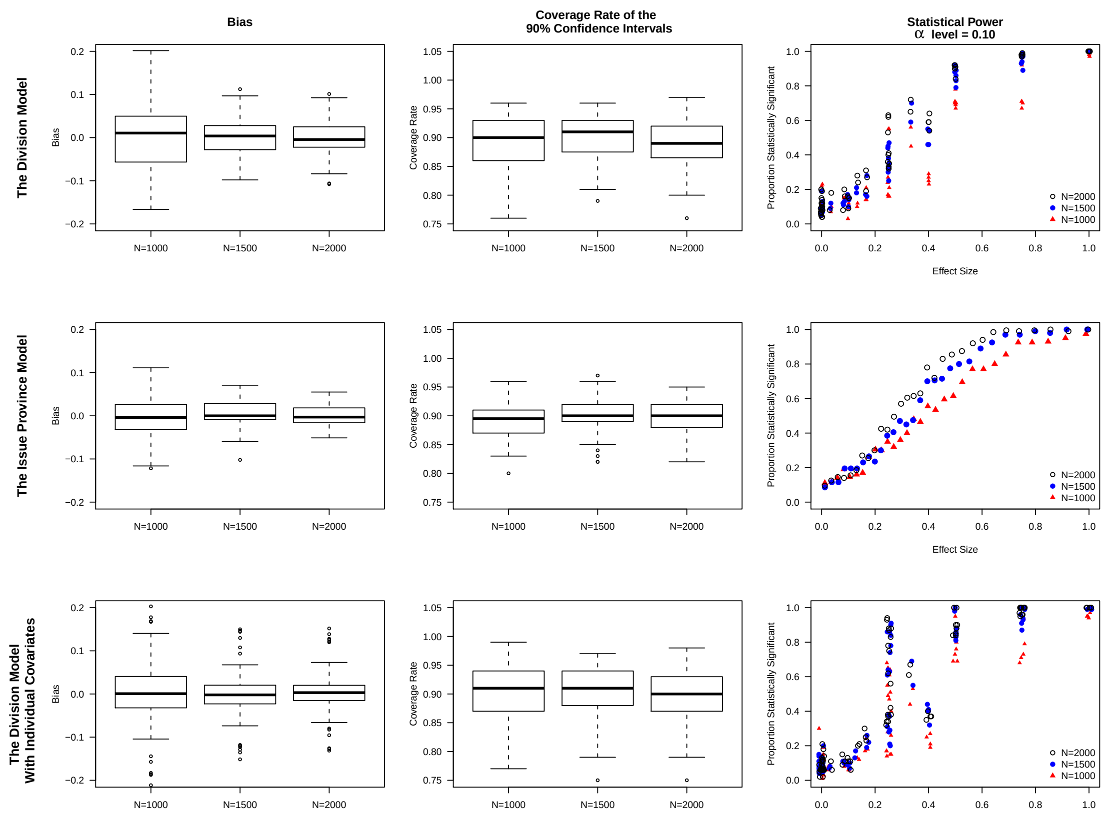
Model for the Province Level Issue Ownership

- 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

- Survey methodology to study sensitive questions
- Endorsement Experiments
  - Most indirect questioning
  - Applicability limited to measuring support
  - Analysis based on the ideal points framework
  - Multilevel modeling to efficient estimation of spatial patterns
- Design considerations:
  - Too many groups $\Longrightarrow$ loss of efficiency
  - Policy positions should not be well-known
  - Response distribution should not be skewed
  - Policies should belong to a single dimension