Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan

Kosuke Imai
Department of Politics
Princeton University

Joint work with Will Bullock and Jacob Shapiro

May 13, 2011
Motivation

- Survey is used widely in social sciences
- Validity of survey depends on the accuracy of self-reports
- **Sensitive questions** $\implies$ 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 Methodologies for Sensitive Questions

- **Randomized Response Technique**
  - Most extensively studied
  - Use *randomization* to protect privacy
  - Difficulties: logistics, lack of understanding among respondents

- **List Experiments** *(item count technique)*
  - Use *aggregation* to protect privacy
  - New *multivariate regression analysis* method
  - New methods to detect and correct failures (joint with G. Blair)
  - Difficulties: design effects, ceiling and floor effects

- **Endorsement Experiments**
  - Use randomized *endorsements* to measure support levels
  - Develop a measurement model based on *item response theory*
  - Difficulties: interpretation, need for modeling
  - Applications:
    1. Pakistanis’ support for Islamic militant groups
    2. Afghanis’ support for Taliban and ISAF (joint with J. Lyall)
    3. Nigerians’ support for insurgents (joint with G. Blair)
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 in April 2009
- Four militant 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?
  (1) A great deal (2) A lot (3) A moderate amount (4) A little (5) Not at all

- 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?
  (1) A great deal (2) A lot (3) A moderate amount (4) A little (5) Not at all
Distribution of Responses

- Polio Vaccinations
- Curriculum Reform
- FCR Reforms
- Durand Line

Regions: Punjab, Sindh, NWFP, Balochistan

Groups: Firqavarana, Tanzeems, Al–Qaida, Afghan Taliban, Pakistani militant groups in Kashmir, Control Group

Responses: Not At All, A Little, A Moderate Amount, A Lot, A Great Deal

Kosuke Imai (Princeton)
Endorsement Experiments
NEMP (NYU)
Methodological Challenges and Proposed Solutions

1. How to combine responses from multiple questions?
   $$\rightarrow$$ item response theory

2. How to recoup loss of statistical efficiency?
   $$\rightarrow$$ hierarchical modeling

3. How to interpret the “support”?
   $$\rightarrow$$ policy vs. valence

4. How to select policy questions?
   - Policies should belong to a single dimension
   - Respondents should not have strong views
   - Should one use well-known policies?:
     Statistical and substantive tradeoffs
Endorsement Experiments Framework

- $N$ 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 (Clinton, Jackman, and Rivers):
  \[
  U_i(\zeta_{j1}, k) = -\|x_i + s^*_{ijk} - \zeta_{j1}\|^2 + \eta_{ij}
  \]
  \[
  U_i(\zeta_{j0}, k) = -\|x_i + s^*_{ijk} - \zeta_{j0}\|^2 + \nu_{ij}
  \]
  $x_i$ is the ideal point and $s^*_{ijk}$ is the “influence” of endorsement

- The statistical model (item response theory):
  \[
  \Pr(Y_{ij} = 1 \mid T_{ij} = k) = \Pr(Y_{ij}(k) = 1)
  \]
  \[
  = \Pr(U_i(\zeta_{j1}, k) > U_i(\zeta_{j0}, k))
  \]
  \[
  = \Pr(\alpha_j + \beta_j(x_i + s^*_{ijk}) > \epsilon_{ij})
  \]

- Support level: greater support $\iff$ greater prob. of $Y_{ij} = 1$

  \[
  s_{ijk} = \begin{cases} 
  s^*_{ijk} & \text{if } \beta_j \geq 0 \\
  -s^*_{ijk} & \text{otherwise}
  \end{cases}
  \]
Hierarchical modeling:

\[ x_i \sim \mathcal{N}(Z_i^\top \delta, \sigma_x^2) \]

\[ s_{ijk} \sim \mathcal{N}(Z_i^\top \lambda_{jk}, \omega_{jk}^2) \]

\[ \lambda_{jk} \sim \mathcal{N}(\theta_k, \Phi_k) \]

“Noninformative” hyper prior on \((\alpha_j, \beta_j, \delta, \theta_k, \omega_{jk}^2, \Phi_k)\)

Interpretation:

- spacial model vs. factor analysis
- policy vs. valence aspects of support
Quantities of Interest and Model Fitting

- **Average support** level for each militant group $k$

  \[
  \tau_{jk}(Z_i) = Z_i^\top \lambda_{jk} \quad \text{for each policy } j
  \]
  \[
  \kappa_k(Z_i) = Z_i^\top \theta_k \quad \text{averaging over all policies}
  \]

- 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:

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

$$
S_{ijk} \overset{\text{indep.}}{\sim} \mathcal{N}(\lambda_{k, \text{division}[i]}, \omega_k^2)
$$

$$
\delta_{\text{division}[i]} \overset{\text{indep.}}{\sim} \mathcal{N}(\mu_{\text{province}[i]}, \sigma^2_{\text{province}[i]})
$$

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

- Averaging over policies
- Partial pooling across divisions within each province
Model with Individual Covariates

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

$$
\begin{align*}
    x_i & \sim \mathcal{N}(\delta_{\text{division}[i]} + Z_i^T \delta Z, 1) \\
    s_{ijk} & \sim \mathcal{N}(\lambda_{k,\text{division}[i]} + Z_i^T \lambda_k Z, \omega_k^2) \\
    \delta_{\text{division}[i]} & \sim \mathcal{N}(\mu_{\text{province}[i]}, \sigma_{\text{province}[i]}^2) \\
    \lambda_{k,\text{division}[i]} & \sim \mathcal{N}(\theta_{k,\text{province}[i]}, \Phi_{k,\text{province}[i]})
\end{align*}
$$

- 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

Kosuke Imai (Princeton)  Endorsement Experiments  NEMP (NYU)
Some Non-Causal Observations

- Comparison with the “knowledge on the ground”
  - Greatest support in Punjab: consistent
  - Support in Gujranwala but not in Bahawalpur: surprising (US AID?)

- Least tolerant where senior leadership resides
  - Hazara for Al-Qa’ida
  - Quetta and Zhob for Taliban

- Least support where many terrorist attacks before April 2009
  - Hazara, Kohat, Nasirabad, Peshawar, and Quetta all suffered from attacks in early 2009
  - Data on “politically motivated violence” from March 2008 through March 2009 (National Counterterrorism Center’s Worldwide Incident Tracking System)
Association between Support and Violence

- **Strong negative association**

  - Pakistani militant groups in Kashmir
    - Correlation: $-0.574$
  - Militants fighting in Afghanistan
    - Correlation: $-0.594$
  - Al–Qaida
    - Correlation: $-0.468$
  - Firqavarana Tanzeems
    - Correlation: $-0.414$

- **Weaker association for the standard ordered probit model**
  (division dummies, treatment variables, their interactions)

  - Pakistani militant groups in Kashmir
    - Correlation: $-0.061$
  - Militants fighting in Afghanistan
    - Correlation: $-0.365$
  - Al–Qaida
    - Correlation: $0.021$
  - Firqavarana Tanzeems
    - Correlation: $-0.166$
No strong relationship between:
- ideology and violence
- ideology and support

**Division–Level Estimated Ideal Point**

- Number of Incidents
  - Correlation: $\text{correlation} = -0.040$

- Division–Level Average Estimated Support for Militant Groups
  - Correlation: $\text{correlation} = 0.087$
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

The Division Model

With Individual Covariates

Bias

Coverage Rate of the 90% Confidence Intervals

Statistical Power

Effect Size

Proportion Statistically Significant

α level = 0.10

Bias Coverage Rate of the 90% Confidence Intervals

Statistical Power

α level = 0.10

Kosuke Imai (Princeton)

Endorsement Experiments

NEMP (NYU)
Monte Carlo Evidence with Varying Sample Size

The Division Model

Bias

Coverage Rate of the 90% Confidence Intervals

Statistical Power

α level = 0.10

Bias

Coverage Rate

Proportion Statistically Significant

Effect Size

Bias

Coverage Rate

Proportion Statistically Significant

Effect Size

Effect Size

Effect Size

Effect Size

Effect Size
Concluding Remarks

- Survey methodology to study sensitive questions

- **Endorsement Experiments**
  - Most indirect form of questioning
  - Applicability limited to measuring support
  - Analysis based on the item response framework
  - Multilevel modeling to efficient estimation of spatial patterns

- **Design considerations:**
  - Policies should belong to a single dimension
  - Respondents should not have strong opinion
  - Separating policy and valence aspects of support
  - Statistical vs. substantive tradeoffs
  - Could measure policy positions and political knowledge separately

- JAGS code available at the dataverse