

Political Methodology: Applied Statistics in Political Science

Kosuke Imai

Department of Politics

Statistics & Machine Learning @ Princeton Symposium

February 18, 2011

Political Methodology

- Applied statistics in political science
- Relatively young but fast growing field:
 - The 1st annual summer meeting in 1984
 - The 28th annual summer meeting at Princeton this summer
 - The 1st issue of *Political Analysis* published in 1989
 - The most cited journal among over 100 political science journals
- Influence from many other fields
- Examples:
 - Econometrics: instrumental variables methods
 - Psychometrics: item response theory
 - Biostatistics: survival analysis
 - Computer science: analysis of text and speech

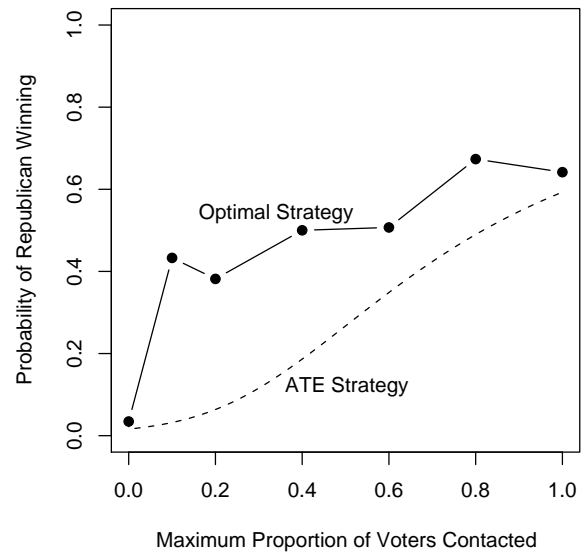
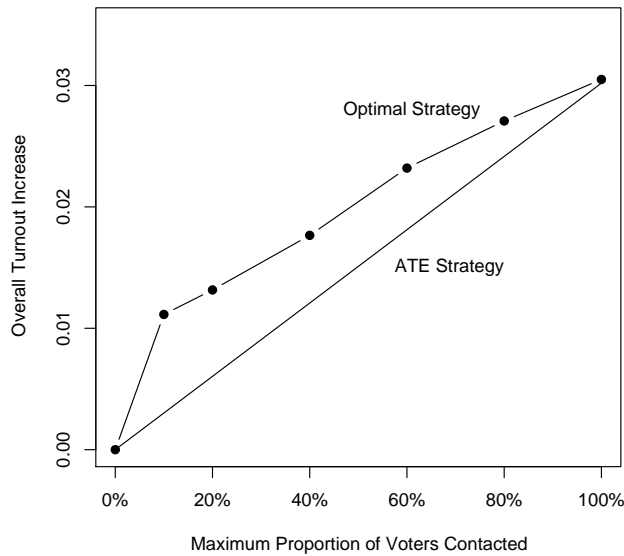
Current Research Projects

- ① Program Evaluation
 - Mexican universal health care program (a.k.a. Seguro Popular)
 - Nigerian conditional oil-revenue transfer program
- ② Statistical Analysis of Causal Mechanisms
 - How, not just whether, does treatment causally affects outcome?
 - Causal mediation analysis, natural direct and indirect effects
 - Identification, inference, sensitivity analysis, experimental designs
- ③ Estimation of Treatment Effect Heterogeneity
 - Which treatment (combination of treatments) works best for whom?
 - Qualitative treatment-covariate/treatment-treatment interactions
 - Use of machine learning methods
- ④ Survey Methodology for Asking Sensitive Questions
 - How to elicit truthful answers to sensitive survey questions?
 - Item count technique (list experiments), endorsement experiments
 - Measuring support for militant groups in Afghanistan and Pakistan

Estimation of Treatment Effect Heterogeneity

- Motivating Application: Optimal Get-out-the-vote Campaigns
 - Non-partisan: maximize turnout
 - Partisan: maximize probability of winning
- Numerous GOTV field experiments with various mobilization strategies
 - Modes: phone, personal visit, postcard, text message, etc.
 - Messages: civic duty, close election, social pressure, etc.
- Question: Which mobilization strategy (combination of strategies) is effective for which voter?

Initial Results based on Classification Trees



- Challenge: Treatment-covariate interactions tend to be overwhelmed by covariate main effects

Development of Alternative Methodology

- Basic problems:
 - 1 Variable selection: finding qualitative interactions
 - 2 Subset selection: finding “marginal” voters
- Support Vector Machine with two separate LASSO constraints:

$$\hat{y}_i = \underbrace{X_i^T}_{\text{other effects}} \beta + \underbrace{Z_i^T}_{\text{interactions}} \gamma$$

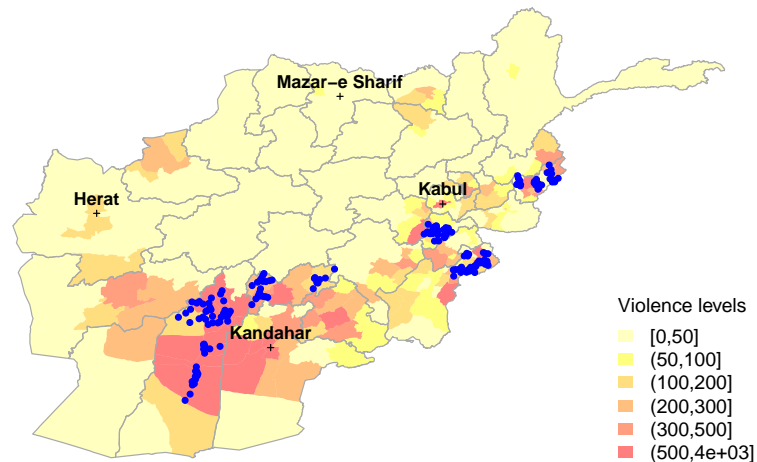
with the following loss function

$$\underbrace{\frac{1}{n} \sum_{i=1}^n |1 - y_i \hat{y}_i|}_{\text{subset selection}} + \underbrace{\lambda_x \sum_{j=1}^k |\beta_j| + \lambda_z \sum_{j=1}^m |\gamma_j|}_{\text{variable selection}} \text{ where } y_i \in \{-1, 1\}$$

- Development of optimization algorithm
- Comparison with Classification Trees, BART, and Boosting

Survey Methodology for Sensitive Questions

- Political scientists use surveys to study sensitive issues such as racial prejudice and corruption
- Direct questioning \implies social desirability bias and nonresponse
- Application in progress: Measuring citizens' support for foreign forces and Taliban in Afghanistan
- Direct questioning \implies you will get lies, nonresponse, and killed



Item Count Technique

- Use aggregation to protect privacy
- 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.

How many, if any, of these things upset you?

Item Count Technique

- Use aggregation to protect privacy
- Randomize the sample into the “treatment” and “control” groups
- The script for the **treatment** group:

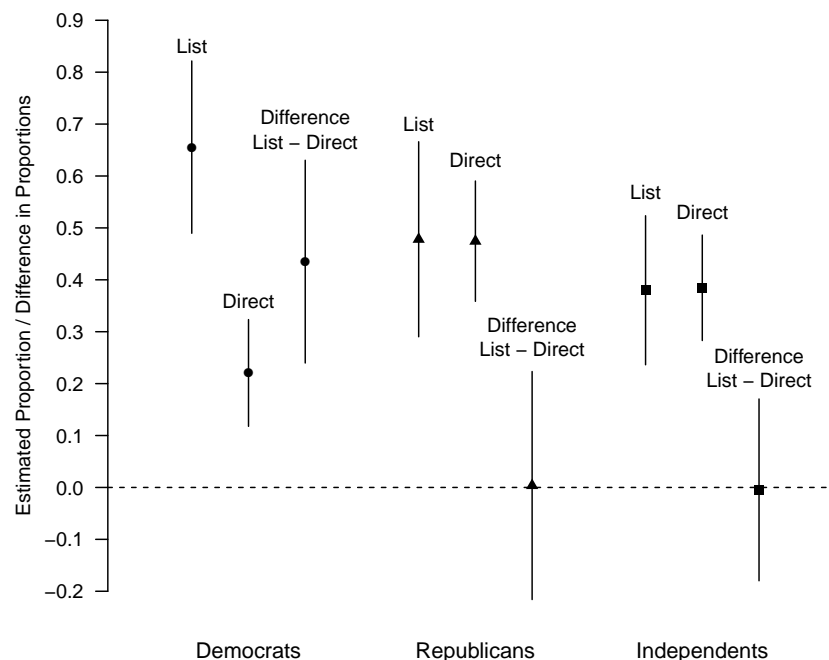
Now I’m going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. (I don’t want to know which ones, just how many.)

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

How many, if any, of these things upset you?

Comparison of Direct and Indirect Questioning

"black leaders asking the government for affirmative action"



Methodological Development and Future Agenda

- Assumptions:
 - ① No Design Effect: Addition of sensitive item does not change responses to control items
 - ② No Liar: Respondents provide truthful response to sensitive item
- What we have developed so far:
 - ① multivariate regression analysis methods
 - ② statistical tests to detect violations of the assumptions
 - ③ statistical methods to model deviations from the assumptions
 - ④ R package that implements these methods
- Next steps:
 - ① extension to a hierarchical model
 - ② spatial pattern of support for Taliban and foreign forces

About Us

- Most political scientists analyze data but few focuses on methodological research
- Marc Ratkovic:
 - Visiting Ph.D. student from Wisconsin finishing up Ph.D.
 - Soon to be a postdoctoral fellow at Princeton
 - Research interests: high-dimensional problems in political science
- Teppei Yamamoto:
 - 5th year graduate student finishing up Ph.D.
 - Soon to be an assistant professor at MIT
 - Research interests: causal inference, modeling of election data
- Where we are: the ground floor of Corwin
- Weekly political methodology seminar: Friday noon in Corwin 127