

Statistical Analysis of List Experiments

Kosuke Imai

Princeton University

Joint work with Graeme Blair

October 8, 2010

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, fraud, support for militant groups
- Lies and non-responses

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

List Experiments and Project Overview

- **List Experiments** (Raghavarao and Federer, 1979)
 - Also known as total block response, item count technique, and unmatched count technique
 - Use aggregation to protect privacy
 - An alternative to randomized response technique
- Goals of this project:
 - Enable *multivariate regression analysis*
 - Develop statistical tests to detect *failures of list experiments*
 - Adjust for deviations from the standard list experiment assumption

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.

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.**

Standard Analysis

- Assumptions:
 - ① **No Design Effect:** The inclusion of the sensitive item does not affect answers to non-sensitive items
 - ② **No Liar:** Answers about the sensitive item are truthful
- Difference-in-means estimator:

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

- Advantages: straightforward, unbiased
- Disadvantages:
 - Inefficient
 - Difficult to explore multivariate relationship
- Need for multivariate regression analysis
- No existing method allows for multivariate regression analysis

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 $\in \{0, 1, \dots, J\}$
 - $Z_{i,J+1}$: truthful answer to the sensitive item $\in \{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 3$)

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* 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 $\{0, 1, \dots, J\}$
 - $Z_{i,J+1}$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 3$)

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* 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 $\{0, 1, \dots, J\}$
 - $Z_{i,J+1}$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $(2 \times J)$ types
- Example: two non-sensitive items ($J = 3$)

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* is identified:

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

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

The Maximum Likelihood (ML) Estimator

- Model for sensitive item as before: e.g., logistic regression

$$\Pr(Z_{i,J+1} = 1 \mid X_i = x) = \text{logit}^{-1}(x^\top \delta)$$

- Model for non-sensitive item given the response to sensitive item: e.g., binomial or beta-binomial regression

$$\Pr(Y_i(0) = y \mid X_i = x, Z_{i,J+1} = z) = \mathcal{J} \times \text{logit}^{-1}(x^\top \psi_z)$$

- Difficult to maximize the resulting likelihood function
- Develop the EM algorithm for reliable estimation

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**

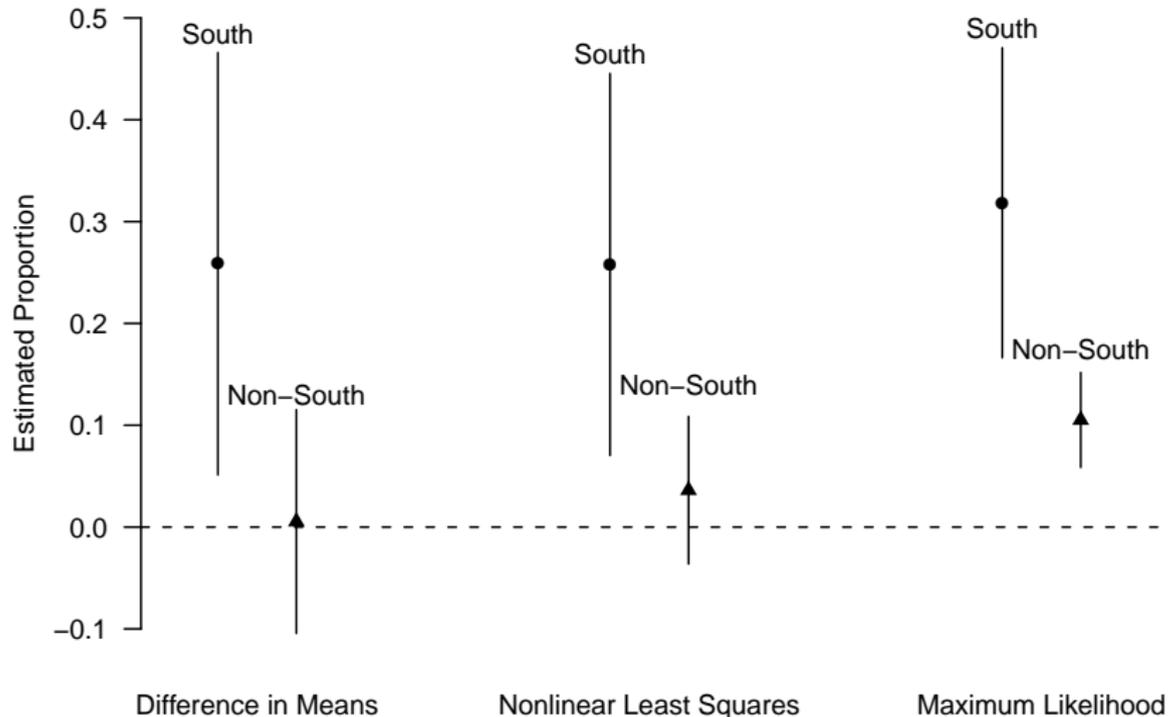
Results of the Multivariate Analysis

- 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

Variables	Nonlinear Least Squares		Maximum Likelihood			
	est.	s.e.	Constrained		Unconstrained	
	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

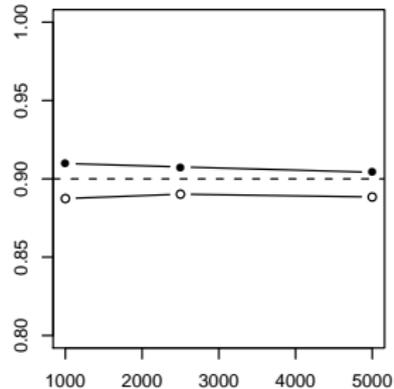
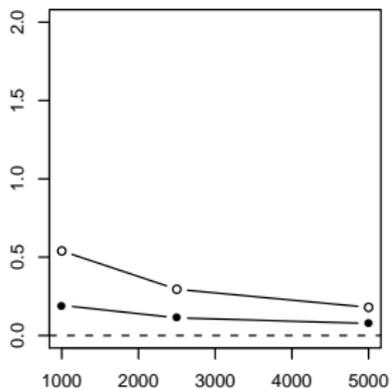
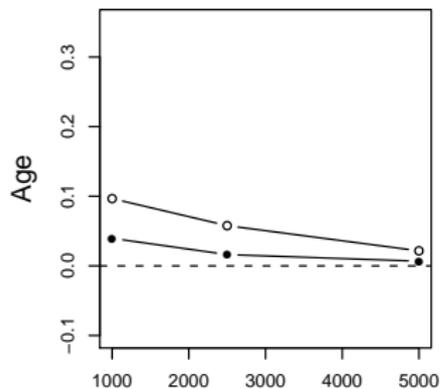
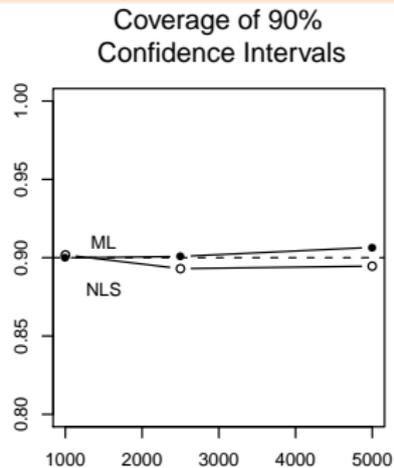
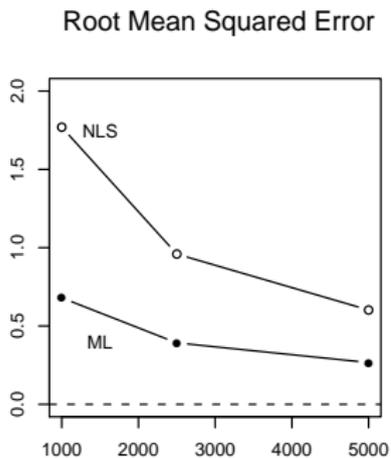
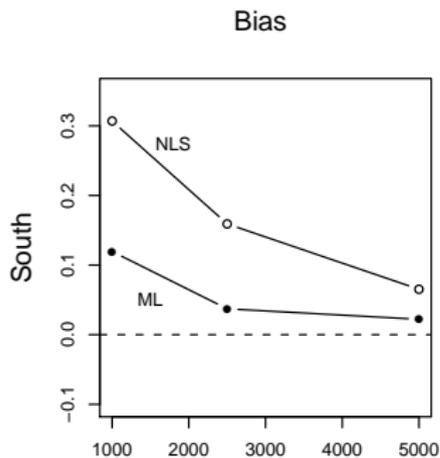
- The original conclusion is supported
- Standard errors are much smaller for ML estimator

Estimated Proportion of Prejudiced Whites



- Regression adjustments and MLE yield more efficient estimates

Simulation Evidence



When Can List Experiments Fail?

- Recall the two assumptions:
 - ① **No Design Effect:** The inclusion of the sensitive item does not affect answers to non-sensitive items
 - ② **No Liar:** Answers about the sensitive item are truthful
- Design Effect:
 - Respondents evaluate non-sensitive items relative to the sensitive item
- Lies:
 - Ceiling effect: too many yeses for non-sensitive items
 - Floor effect: too many noes for non-sensitive items
- Both types of failures are difficult to detect
- Importance of pilot studies: ask non-sensitive and sensitive items separately and compare the responses
- Question: Can these failures be addressed statistically?

Hypothesis Tests for Design Effects

- Under the assumption of no design effect, we have a non-negative proportion for each respondent type

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

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

- Null hypothesis: ALL of these proportions are non-negative
- Alternative hypothesis: At least one is negative
- Two multivariate one-sided tests with Bonferroni correction
- The proposed statistical test adjusts for multiple testing problem
- Failure to reject the null may arise from the lack of power

The Racial Prejudice Data Revisited

- Did the negative proportion arise by chance?

y value	Observed Data				Estimated Proportion of Respondent Types			
	Control		Treatment		$\hat{\pi}_{y0}$	s.e.	$\hat{\pi}_{y1}$	s.e.
0	8	1.4%	19	3.0%	3.0%	0.7	-1.7%	0.8
1	132	22.4	123	19.7	21.4	1.7	1.0	2.4
2	222	37.7	229	36.7	35.7	2.6	2.0	2.8
3	227	38.5	219	35.1	33.1	2.2	5.4	0.9
4			34	5.4				
Total	589		624		93.2		6.8	

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 there is no design effect
- Proposed strategy: model ceiling and/or floor effects under an additional assumption
- **Identification assumption**: conditional independence between items given covariates
- ML estimation can be extended to this situation

Statistical Analysis of the Modified Design

- Corstange (2009) proposes to ask each non-sensitive question directly in the control group
- Advantage: more efficient than the standard design
- Disadvantage: possible design effect
- Extend the NLS and ML estimators to this design
- Correct a mistake in the formulation of the likelihood
- Develop the EM algorithm for reliable estimation
- A simulation study suggests the proposed estimators outperform the existing estimator in terms of bias, efficiency, and coverage

Concluding Remarks

- List experiments: survey technique for eliciting truthful responses
- Advantages: easy to use, easy to understand
- Disadvantages:
 - ① inefficient
 - ② difficult to analyze multivariate relationship
 - ③ the assumptions may be violated
- The importance of design: choice of non-sensitive items
- Our propose methods partially overcome the difficulties
 - multivariate regression analysis for efficient analysis
 - exploration of multivariate relationship
 - statistical tests for detecting design effect
 - modeling ceiling and floor effects

- **PAPERS:**

- ① Imai. “Statistical Inference for the Item Count Technique.”
- ② Blair and Imai. “Statistical Analysis of List Experiments.”

- **SOFTWARE:** R package

Blair, Graeme, and Kosuke Imai. `list`: Multivariate Statistical Analysis for the Item Count Technique.

- **PROJECT WEBSITE:**

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