

Statistical Analysis of List Experiments

Graeme Blair

Kosuke Imai

Princeton University

March 31, 2011

Methodological Challenges

- List experiment is becoming popular:
Kuklinski et al., 1997a,b; Sniderman and Carmines, 1997; Gilens et al., 1998; Kane et al., 2004; Tsuchiya et al., 2007; Streb et al., 2008; Corstange, 2009; Flavin and Keane, 2010; Glynn, 2010; Gonzalez-Ocantos et al., 2010; Holbrook and Krosnick, 2010; Janus, 2010; Redlawsk et al., 2010; Coutts and Jann, 2011
- Standard practice: Use difference-in-means to estimate the proportion of those who answer yes to sensitive item
- Getting more out of list experiments:
 - ① Who are more likely to answer yes?
 - ② Who are answering differently to direct and indirect questioning?
 - ③ Can we study multiple sensitive items in one survey?
 - ④ Can we detect failures of list experiments?
 - ⑤ Can we correct violations of key assumptions?
- Recoup the efficiency loss due to indirect questioning

Overview of the Project

- Goals:
 - ① Develop *multivariate regression analysis* methodology
 - ② Develop statistical tests to detect *failures of list experiments*
 - ③ Develop methods to correct deviations from key assumption
 - ④ Develop open-source software to implement the proposed methods
 - ⑤ Applications in Afghanistan (joint work with J. Lyall) and Nigeria
- References:
 - ① Imai, K. (in-press) “Multivariate Regression Analysis for the Item Count Technique.” *Journal of the American Statistical Association*
 - ② Blair, G. and K. Imai. “Statistical Analysis of List Experiments.”
 - ③ Blair, G. and K. Imai. `list`: Statistical Methods for the Item Count Technique and List Experiments available at <http://cran.r-project.org/package=list>

Identification Assumptions

- ① Randomization of the Treatment
- ② **No Design Effect:** The inclusion of the sensitive item does not affect answers to control items
- ③ **No Liars:** Answers about the sensitive item are truthful

Under these assumptions, difference-in-means estimator is unbiased

New Multivariate Regression Estimators

- Notation:
 - J : number of control items
 - N : number of respondents
 - T_i : binary treatment indicator (1 = treatment, 0 = control)
 - X_i : pre-treatment covariates
 - Y_i : outcome variable
- **The nonlinear least squares regression model:**

$$Y_i = \underbrace{f(X_i, \gamma)}_{\text{control items}} + \underbrace{T_i \cdot g(X_i, \delta)}_{\text{sensitive item}} + \epsilon_i$$

- Difference-in-means: no covariate
- Linear model: $f(x, \gamma) = x^\top \gamma$ and $g(x, \delta) = x^\top \delta$
- Logit model: $f(x, \gamma) = J \cdot \text{logit}^{-1}(x^\top \gamma)$ and $g(x, \delta) = \text{logit}^{-1}(x^\top \delta)$
- Two-step estimation with appropriate standard error

The Maximum Likelihood Estimator

- Under assumptions, joint distribution of $(Z_{i,J+1}^*, Y_i(0))$ is identified
- More information can be extracted!
- 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 control items given the response to sensitive item: e.g., binomial or beta-binomial logistic regression

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

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

Empirical Application: Racial Prejudice in the US

- Kuklinski *et al.* (1997 JOP): Southern whites are more prejudiced against blacks than non-southern whites – no “New South”

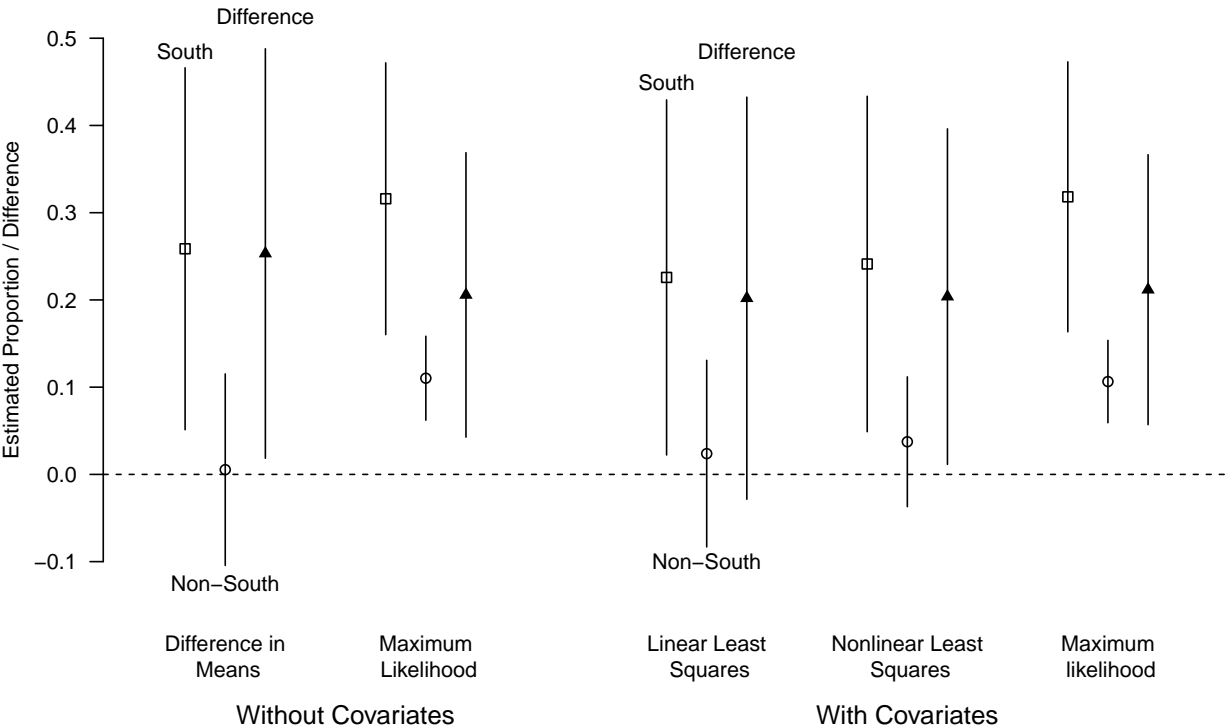
(4) a black family moving next door to you

- 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

- Need for a **multivariate regression analysis**

Estimated Proportion of Prejudiced Whites

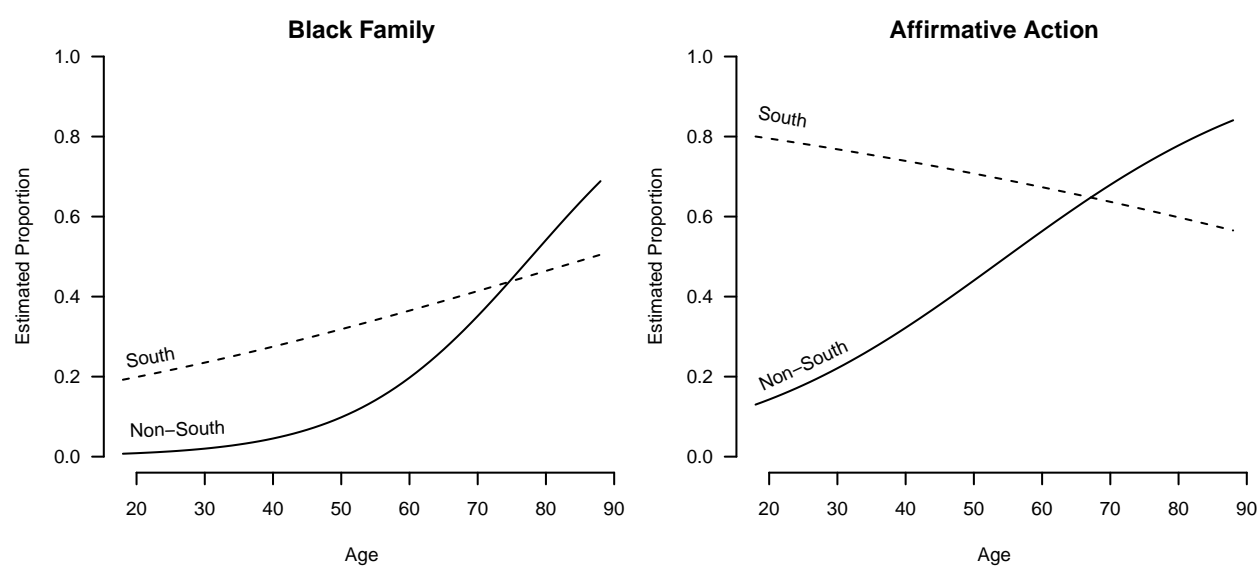


- MLE yields more efficient estimates

Studying Multiple Sensitive Items

- The 1991 National Race and Politics Survey includes another treatment group with the following sensitive item
(4) "black leaders asking the government for affirmative action"
- Use of the same control items permits joint-modeling
- Same assumptions: No Design Effect and No Liars
- Extension to the design with K sensitive items
- How do the patterns of generational changes differ between South and Non-South?
- Original analysis dichotomized the age variable without controlling for other factors

Generational Changes in South and Non-South



- Age is important even after controlling for gender and education
- Gender is not, contradicting with the original analysis

Measuring Social Desirability Bias

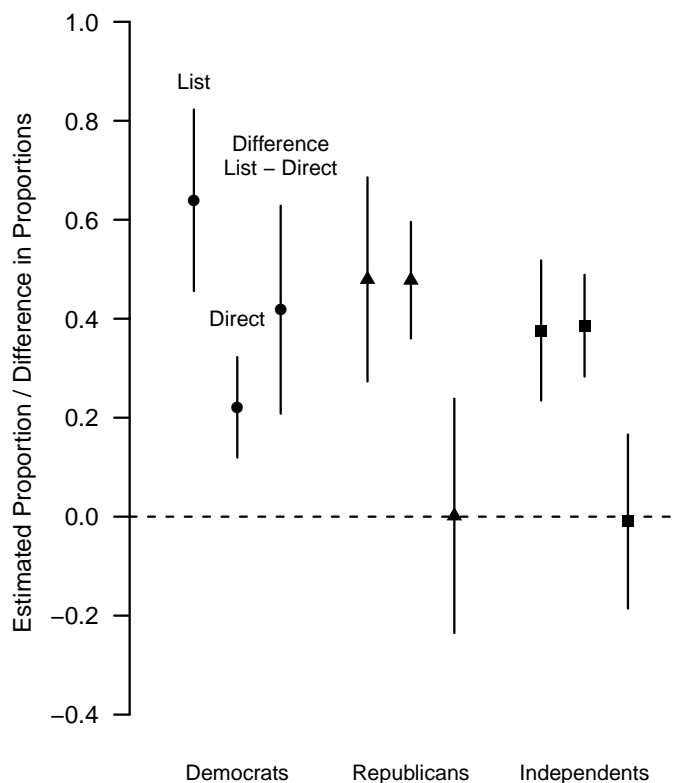
- The 1994 Multi-Investigator Survey (Sniderman et al.) asks list experiment question and later a **direct sensitive question**:

Now I'm going to ask you about another thing that sometimes makes people angry or upset.

Do you get angry or upset when black leaders ask the government for affirmative action?

- Difference between direct and indirect responses
⇒ measure of social desirability bias

Social Desirability Bias by Partisanship



When Can List Experiments Fail?

- **No Design Effect**
- Respondents may evaluate control items relative to sensitive item
- **No Liars**
- Ceiling effect: too many yeses for control items
- Floor effect: too many noes for control items
- Question: Can these failures be addressed statistically?

Hypothesis Test for Detecting List Experiment Failures

- Under the **null hypothesis** of no design effect, we expect proportions of all “types” to be properly estimated
- **Alternative hypothesis**: *At least one is negative*
- Correction for multiple testing
- Estimated proportions of respondent types

Response for control items	Response for sensitive item			
	Yes		No	
	est.	s.e.	est.	s.e.
0	3.0%	0.7	-1.7%	0.8
1	21.4	1.7	1.0	2.4
2	35.7	2.6	2.0	2.8
3	33.1	2.2	5.4	0.9

- $p\text{-value} = 0.022$

Modeling Ceiling and Floor Effects

- Proposed strategy: model ceiling and/or floor effects under an additional assumption
- **Identification assumption**: conditional independence between items given covariates
- ML regression estimator can be extended to this situation
- A similar strategy applicable to design effects

Concluding Remarks and Practical Suggestions

- List experiments: easy to use, easy to understand
- Challenges:
 - ① loss of information due to indirect questioning
 - ② difficulty of exploring multivariate relationship
 - ③ potential violation of assumptions
- Our methods partially overcome the difficulties
- Suggestions for analysis:
 - ① Estimate proportions of types and test design effects
 - ② Conduct multivariate regression analyses
 - ③ Investigate the robustness of findings to ceiling and floor effects
- Suggestions for design:
 - ① Select control items to avoid skewed response distribution
 - ② Avoid control items that are ambiguous and generate weak opinion
 - ③ Conduct a pilot study and maximize statistical power