Statistical Analysis of List Experiments

Graeme Blair Kosuke Imai

Princeton University

December 17, 2010

Motivation

- Surveys are used widely in social sciences
- Validity of surveys depends on the accuracy of self-reports
- Sensitive questions $\Longrightarrow$ 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

Project Overview

- **List Experiments**
  - Also known as total block response, item count technique, and unmatched count technique
  - Use aggregation to protect privacy
  - An alternative to randomized response technique
  - Little methodological work

- Goals of this project:
  1. Formalize the key identification assumptions
  2. Develop methods for multivariate regression analysis
  3. Develop a statistical test to detect failures of list experiments
  4. Develop methods to adjust for deviations from the assumptions
  5. Develop software to implement all suggestions
  6. Extend the methods to measure spacial variation of citizens' support for militant groups and foreign forces

Project Reference

- Papers:
  1. Imai. "Statistical Inference for the Item Count Technique."
  2. Blair and Imai. "Statistical Analysis of List Experiments."

- Software:
  Blair and Imai. `list: Multivariate Statistical Analysis for the Item Count Technique`. R package

- Applications (in Progress):
  1. Measuring support for foreign forces and Taliban in Afghanistan (joint with Jason Lyall)
  2. Measuring support for insurgent groups in the Niger Delta

- Project Website:

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

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.


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

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.

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

Notation and Setup

- J: number of non-sensitive items
- N: number of respondents
- T_i: binary treatment indicator (1 = treatment, 0 = control)

- Potential outcomes notation
- $Z_{ij}(t)$: potential response to the jth non-sensitive item under treatment status $T_i = t$ for $j = 1, \ldots, J$ and $t = 0, 1$
- $Z_{i,J+1}(t)$: potential response to the sensitive item under treatment status $T_i = t$ where $Z_{i,J+1}(0)$ represents truthful answer

- $Y_i(0) = \sum_{j=1}^{J} Z_{ij}(0)$: potential response under control condition
- $Y_i(1) = \sum_{j=1}^{J+1} Z_{ij}(1)$: potential response under treatment condition
- $Y_i = Y_i(T_i)$: observed response

Identification Assumptions

1. No Design Effect: The inclusion of the sensitive item does not affect answers to non-sensitive items

$$\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1)$$

2. No Liars: Answers about the sensitive item are truthful

$$Z_{i,J+1}(0) = Z_{i,J+1}(1)$$

Limitations to the Standard Techniques

- Difference-in-means estimator:

$$\hat{\tau} = \text{average of the treated} - \text{average of the control}$$

- Straightforward and unbiased under the above assumptions
- But, potentially inefficient
- Difficult to explore multivariate relationship
- 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:

$$Y_i = \underbrace{f(X_i, \gamma)}_{non-sensitive} + T_i \times \underbrace{g(X_i, \delta)}_{sensitive} + \epsilon_i$$

  - X_i: covariates
  - $f(x, \gamma)$: model for non-sensitive items, e.g., $J \times \mathrm{logit}^{-1}(x^\top \gamma)$
  - $g(x, \delta)$: model for sensitive item, e.g., $\mathrm{logit}^{-1}(x^\top \delta)$
- Two-step estimation procedure:
  1. Fit $f(x, \gamma)$ to the control group via NLS and obtain $\hat{\gamma}$
  2. Fit $g(x, \delta)$ to the treatment group via NLS after subtracting $f(X_i, \hat{\gamma})$ from Y_i and obtain $\hat{\delta}$
- Standard errors via the generalized method of moments
- With no covariates, it reduces to the difference-in-means estimator

Extracting More Information from the Data

- Define a "type" of each respondent by $(Y_i(0), Z_{i,J+1}(0))$
  - $Y_i(0)$: total number of yes for non-sensitive items $\in \{0, 1, \ldots, J\}$
  - $Z_{i,J+1}(0)$: truthful answer to the sensitive item $\in \{0, 1\}$
- A total of $2 \times (J + 1)$ 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 $(Y_i(0), Z_{i,J+1}(0))$
  - $Y_i(0)$: total number of yes for non-sensitive items $\{0, 1, \ldots, J\}$
  - $Z_{i,J+1}(0)$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $2 \times (J + 1)$ 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 $(Y_i(0), Z_{i,J+1}(0))$
  - $Y_i(0)$: total number of yes for non-sensitive items $\{0, 1, \dots, J\}$
  - $Z_{i,J+1}(0)$: truthful answer to the sensitive item $\{0, 1\}$
- A total of $2 \times (J + 1)$ 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

$$\Pr(Z_{i,J+1}(0) = 1 \mid X_i = x) \;=\; \text{logit}^{-1}(x^\top \delta)$$

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

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

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

The Likelihood Function

- $g(x, \delta) = \Pr(Z_{i,J+1}(0) = 1 \mid X_i = x)$
- $h_z(y; x, \psi_z) = \Pr(Y_i(0) = y \mid X_i = x, Z_{i,J+1}(0) = z)$
- The likelihood function consists of mixtures:

$$\prod_{i \in \mathcal{J}(1,0)} (1 - g(X_i, \delta)) h_0(0; X_i, \psi_0) \prod_{i \in \mathcal{J}(1,J+1)} g(X_i, \delta) h_1(J; X_i, \psi_1)$$

$$\times \prod_{y=1}^{J} \prod_{i \in \mathcal{J}(1,y)} \{g(X_i, \delta) h_1(y-1; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

$$\times \prod_{y=0}^{J} \prod_{i \in \mathcal{J}(0,y)} \{g(X_i, \delta) h_1(y; X_i, \psi_1) + (1 - g(X_i, \delta)) h_0(y; X_i, \psi_0)\}$$

where $\mathcal{J}(t, y)$ represents a set of respondents with $(T_i, Y_i) = (t, y)$

- Maximizing this function is difficult

Missing Data Framework and the EM Algorithm

- Consider $Z_{i,J+1}(0)$ as missing data
- For some respondents, $Z_{i,J+1}(0)$ is "observed"
- The complete-data likelihood has a simple form:

$$\prod_{i=1}^{N} \left\{ g(X_i,\delta) h_1(Y_i - 1; X_i, \psi_1)^{T_i} h_1(Y_i; X_i, \psi_1)^{1-T_i} \right\}^{Z_{i,J+1}(0)} \times \{(1 - g(X_i,\delta)) h_0(Y_i; X_i, \psi_0)\}^{1-Z_{i,J+1}(0)}$$

- The EM algorithm: only separate optimization of $g(x,\delta)$ and $h_z(y; x, \psi_z)$ is required
  - weighted logistic regression
  - weighted binomial logistic regression
- Both can be implemented in standard statistical software

Empirical Application: Racial Prejudice in the US

- Kuklinski, Cobb, and Gilens (1997) analyze the 1991 National Race and Politics survey with the difference-in-means estimator
- Finding: Southern whites are more prejudiced against blacks than non-southern whites – no evidence for the "New South"
- The limitation acknowledged by the authors:
  "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

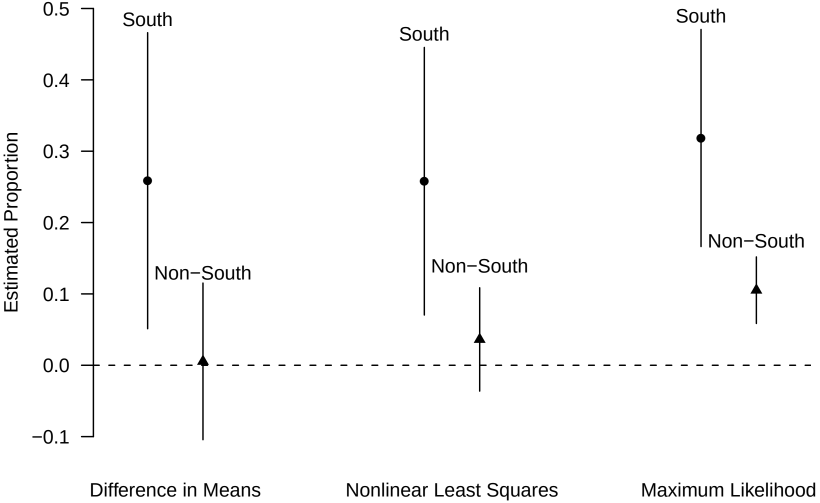
Results of the Multivariate Analysis

- 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

	Nonlinear Least Squares		Maximum Likelihood			
			Constrained		Unconstrained	
Variables	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

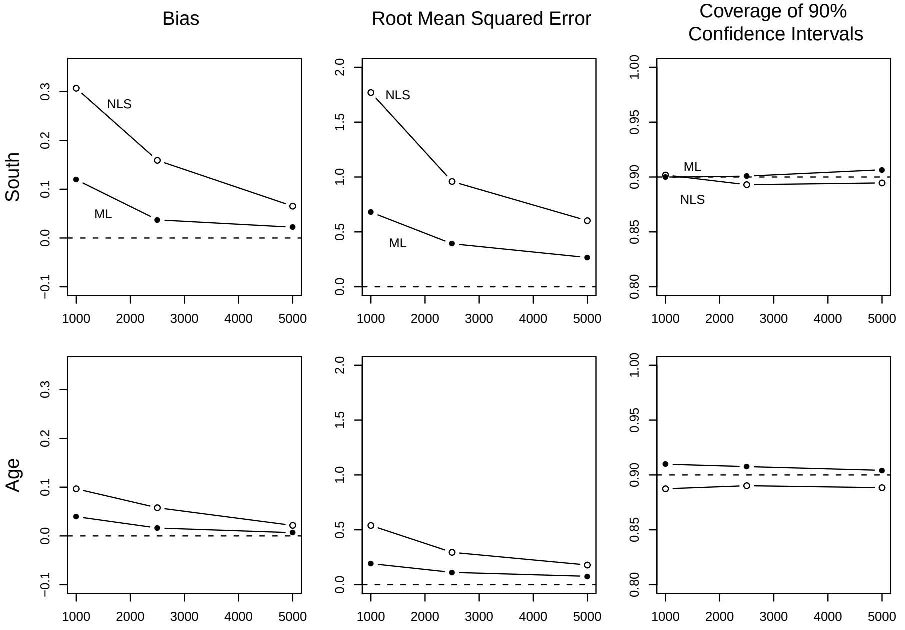
- The original conclusion is supported
- Standard errors are smaller for ML estimator

Estimated Proportion of Prejudiced Whites



- Regression adjustments and MLE yield more efficient estimates

Simulation Evidence



The Design with 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 non-sensitive items permits joint-modeling
- Same assumptions: No Design Effect and No Liars
- Extension to the design with K sensitive items:

$$\begin{aligned} h(y; x, \psi) &= \Pr(Y_i(0) = y \mid X_i = x) \\ g_t(x, y, \delta_{ty}) &= \Pr(Z_{i,J+t}(0) = 1 \mid Y_i(0) = y, X_i = x) \end{aligned}$$

  for each $t = 1, \ldots, K$
- The EM algorithm for the ML estimation

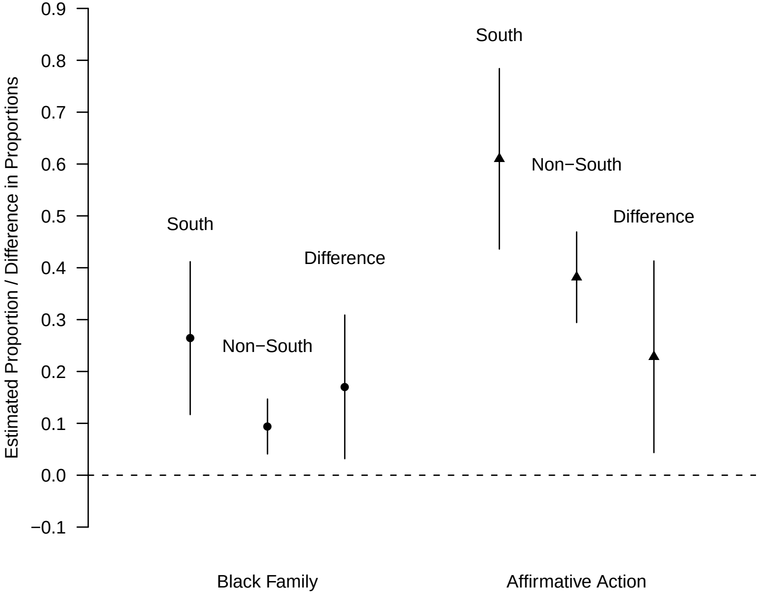
Results of Multivariate Analysis with Joint Modeling

- Multivariate analysis results:

	Sensitive Items				Non-Sensitive Items	
	Black Family		Affirmative Action			
Variables	est.	s.e.	est.	s.e.	est.	s.e.
Intercept	−6.886	1.357	−4.203	1.137	1.296	0.127
Male	0.782	0.524	0.083	0.407	−0.262	0.073
College	−0.356	0.494	−0.445	0.386	−0.543	0.074
Age	0.669	0.169	0.431	0.129	0.025	0.024
South	1.663	0.580	1.332	0.542	−0.228	0.088
$Y_i(0)$	0.424	0.309	0.890	0.248		

- Results are consistent with the "No New-South" finding

Estimated Proportion of Prejudiced Whites



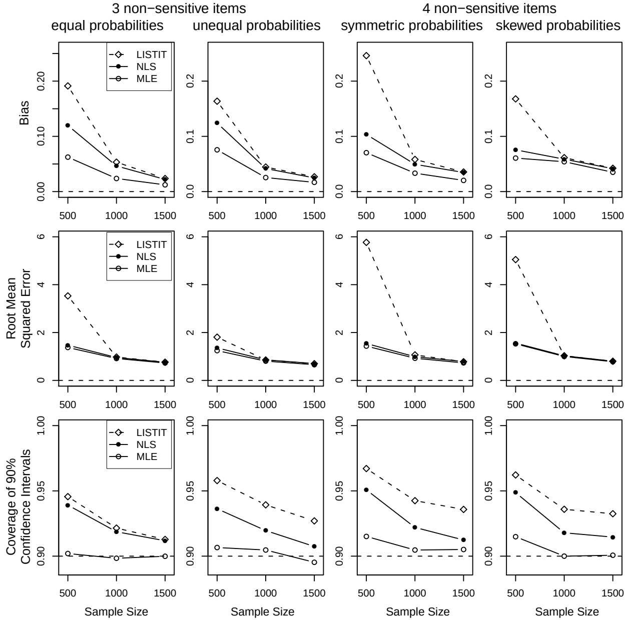
The Design with Non-Sensitive Items Asked Directly

- Motivation: Indirect questioning $\Longrightarrow$ loss of information
- Recoup lost efficiency by asking non-sensitive items directly (for the control group)

- Corstange (2009) proposes a multivariate regression model based on separate logistic regression for each item (LISTIT)
- Concern of design effect raised by Flavin and Keane (2010)

- Likelihood must be based on Poisson-Binomial instead of Binomial
- Develop the EM algorithm for reliable ML estimation
- Also extend the NLS estimator to this design

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 Liars: Answers about the sensitive item are truthful
- Design Effect:
  - Respondents evaluate non-sensitive items relative to 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 choosing non-sensitive items
- Question: Can these failures be addressed statistically?

Hypothesis Test for Detecting List Experiment Failures

- Under the null hypothesis of no design effect and no liars, we

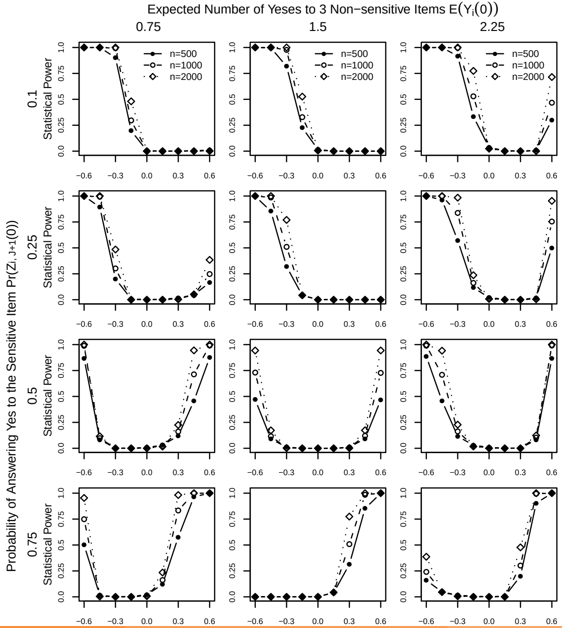
$$\pi_1 = \Pr(\text{type} = (y,1)) = \Pr(Y_i \leq y \mid T_i = 0) - \Pr(Y_i \leq y \mid T_i = 1) \geq 0$$
$$\pi_0 = \Pr(\text{type} = (y,0)) = \Pr(Y_i \leq y \mid T_i = 1) - \Pr(Y_i < y \mid T_i = 0) \geq 0$$

- Alternative hypothesis: *At least* one is negative
- A multivariate one-sided LR test for each $t = 0, 1$

$$\hat{\lambda}_t = \min_{\pi_t} (\hat{\pi}_t - \pi_t)^\top \widehat{\Sigma}_t^{-1} (\hat{\pi}_t - \pi_t), \quad \text{subject to } \pi_t \geq 0,$$

  where $\hat{\lambda}_t$ follows a mixture of χ^2
- Difficult to characterize least favorable values under the joint null
- Bonferroni correction: Reject the joint null if $\min(\hat{p}_0, \hat{p}_1) \leq \alpha/2$
- Failure to reject the null may arise from the lack of power
- We use the Generalized Moment Selection to improve power

Statistical Power of the Proposed Test



The Racial Prejudice Data Revisited

- Let's look at the black family item
- Did the negative proportion arise by chance?

	Observed Data				Estimated Proportion of Respondent Types			
	Control		Treatment					
Response	counts	prop.	counts	prop.	$\hat{\pi}_{y0}$	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	

- Minimum *p*-value: 0.022
- Fail to reject the null with $\alpha = 0.05$
- For the affirmative action item, *p*-value is 0.394

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)

- The above test does not detect these liars 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

Ceiling and Floor Effects for the Affirmative Action Item

Variables	Ceiling Effects Alone est.	Ceiling Effects Alone s.e.	Floor Effects Alone est.	Floor Effects Alone s.e.	Both Ceiling and Floor Effects est.	Both Ceiling and Floor Effects s.e.
Intercept	−1.291	0.558	−1.251	0.501	−1.245	0.502
Age	0.294	0.101	0.314	0.092	0.313	0.092
College	−0.345	0.336	−0.605	0.298	−0.606	0.298
Male	0.038	0.346	−0.088	0.300	−0.088	0.300
South	1.175	0.480	0.682	0.335	0.681	0.335
Prop. of liars						
Ceiling	0.0002	0.0017			0.0002	0.0016
Floor			0.0115	0.0000	0.0115	0.0000

- Essentially no ceiling and floor effects
- Main conclusion for the affirmative action item seems robust

Concluding Remarks

- List experiments as alternative to randomized response method
- Advantages:
  1. some empirical evidence that list experiments "work" better
  2. easy to use, easy to understand
- Disadvantages:
  1. loss of information $\Longrightarrow$ inefficiency
  2. difficult to explore multivariate relationship
  3. identification assumptions may be violated
- Our proposed methods partially overcome the difficulties:
  - multivariate regression analysis
  - statistical test for detecting list experiment failures
  - adjusting for ceiling and floor effects
- The importance of design: choice of non-sensitive items
- Future work:
  - Applications in Afghanistan and Nigeria
  - Extension to hierarchical models for identifying spatial variation