Supplemental Appendix for "List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators"

John S. Ahlquist

July 18, 2017

A Identifying joint proportions

The DiM estimator relies on relatively weak assumptions. Glynn (2013) invokes (by other names) the stronger *no design effects* and *no liars* assumptions in order to characterize the joint distribution of $(Y_i(0), Z_i^*)$, thereby generating estimates of the population proportion which would answer affirmatively to the sensitive item and exactly $j \leq J$ of the control items. It is then a short jump to including covariates.

To see this intuitively, consider a list experiment with $J = 4$ baseline items, as in Table 1.[1] Define $\mathcal{K}(y, z^*)$ as the set of individuals with values $(Y_i(0), Z_i^*)$, i.e. the set that would respond affirmatively to y baseline items (under the no-treatment condition) and with true response of z^* for the sensitive item. Let π_{y1} be the population proportion of people who would answer yes to y control items and the sensitive item. Under the maintained identification assumptions we know that all respondents who answer "0" in the treatment condition–the first cell in the third column of the table–are certainly in $\mathcal{K}(0, 0)$ and all who answer "5" are in $\mathcal{K}(4, 1)$. The remainder of the table describes the other combinations.

Table 1: Respondent types identified under the *no design effect* and *no liars* assumptions for a $J = 4$ list experiment.

y_i	Baseline	Treatment
0	$\mathcal{K}(0,0) \cup \mathcal{K}(0,1)$	$\mathcal{K}(0,0)$
1	$\mathcal{K}(1,0) \cup \mathcal{K}(1,1)$	$\mathcal{K}(1,0) \cup \mathcal{K}(0,1)$
2	$\mathcal{K}(2,0) \cup \mathcal{K}(2,1)$	$\mathcal{K}(2,0) \cup \mathcal{K}(1,1)$
3	$\mathcal{K}(3,0) \cup \mathcal{K}(3,1)$	$\mathcal{K}(3,0) \cup \mathcal{K}(2,1)$
4	$\mathcal{K}(4,0) \cup \mathcal{K}(4,1)$	$\mathcal{K}(4,0) \cup \mathcal{K}(3,1)$
5	$\varnothing$	$\mathcal{K}(4,1)$

[1] See Glynn (2013:appendix D) for a more general formal derivation.

From here we can characterize other quantities, for example $\hat{\pi}_{21}$. By the identification assumptions, the baseline respondents answering "3" or higher are all the baseline respondents in $\mathcal{K}(3,0) \cup \mathcal{K}(3,1) \cup \mathcal{K}(4,0) \cup \mathcal{K}(4,1)$. Similarly eone who answers "3" or higher in the treatment condition are the treatment respondents in $\mathcal{K}(2,1) \cup \mathcal{K}(3,0) \cup \mathcal{K}(3,1) \cup \mathcal{K}(4,0) \cup \mathcal{K}(4,1)$. The disjointness of all these $\mathcal{K}(y, z*)$ sets implies that $|\{y_i : y_i \geq 3, T_i = 0\}|/(N - N_1)$ is an unbiased estimator of $(\pi_{30} + \pi_{31} + \pi_{40} + \pi_{41})$. Similarly $|\{y_i : y_i \geq 3, T_i = 1\}|/N_1$ is an unbiased estimator of $(\pi_{21} + \pi_{30} + \pi_{31} + \pi_{40} + \pi_{41})$. Thus an unbiased estimator of $\hat{\pi}_{21}$ is

$$\hat{\pi}_{21} = |\{y_i : y_i \geq 3, T_i = 1\}|/N_1 - |\{y_i : y_i \geq 3, T_i = 0\}|/(N - N_1)$$

Clearly this exercise can be repeated to estimate any of the π_{y1} quantities.

Summing the $\hat{\pi}_{y1}$ yields the DiM estimator. To see this, Let $N_1^x = |\{i : y_i = x, T_i = 1\}|$ and $(N - N_1)^x = |\{i : y_i = x, T_i = 0\}|$ we can rewrite Equation 3 in the main text as

$$\begin{aligned}
\hat{\tau} &= \frac{1}{N_1}\left[(J+1)N_1^{J+1} + JN_1^J + \ldots + 0N_1^0\right] - \frac{1}{N - N_1}\left[J(N-N_1)^J + \ldots + 0(N-N_1)^0\right] \\
&= \frac{1}{N_1}\sum_{j=0}^{J+1}\sum_{i=j}^{J+1} N_1^i - \frac{1}{N-N_1}\sum_{j=0}^{J}\sum_{i=j}^{J}(N-N_1)^i \\
&= \sum_{k=0}^{J+1}\left[\frac{1}{N_1}\sum_{j=k}^{J+1} N_1^j - \frac{1}{N-N_1}\sum_{j=k}^{J}(N-N_1)^j\right]
\end{aligned}$$

That is, we can calculate the DiM estimator by taking the difference between baseline proportion and treatment proportion of respondents saying at least j and then summing these differences across $j = 0, \ldots, J + 1$

B Additional Detail on the Monte Carlo Experiments

I examine $J = 3$ and $J = 4$ lists. To generate these lists each individual has J binary attributes, denoted $\{C_1, \ldots, C_J\}$, used to generate a hypothetical respondent's values for each of J "control" items. The parameter values for each of the attributes for each list are displayed in table 2.

I investigate two different sets of list structures. The first set of lists generates control items following two of the protocols that Blair and Imai (2012) take from Corstange (2009). In these lists all control items are independent. In the second set—referred to as the "designed" lists—I construct the control item lists to conform to current recommendations for avoiding strategic misrepresentation. In both the $J = 3$ and $J = 4$ designed lists C_1 and C_2 are relatively common in the population but with a moderate negative correlation, following the discussion in Glynn (2013). C_3 and C_4 are relatively low- and high-frequency attributes, respectively, designed to reduce the risk of ceiling and floor effects.

For each of the 2000 Monte Carlo runs I generate a sample of $N = 1000$ "respondents." With equal probability I randomly assign each of the respondents to be in the treatment

Table 2: Parameters for the control item lists in the Monte Carlo experiments.

	Blair-Imai			Designed		
	$J=3$	$J=4$	Correlation	$J=3$	$J=4$	Correlation
C_1	0.50	1/6	independent	0.50	0.50	$\text{cor}(C_1, C_2) = -0.6$
C_2	0.50	1/2	independent	0.50	0.50	$\text{cor}(C_1, C_2) = -0.6$
C_3	0.50	2/3	independent	0.15	0.15	independent
C_4		2/3	independent		0.85	independent

group or the control group, denoted by binary variable T_i. For each of the respondents we then calculate the the error-free observed outcome for $k \in \{L, M, H\}$ as

$$\begin{aligned} y_i^k \mid T^i = 0 &= \sum_{j=1}^{J} c_{ij} \\ y_i^k \mid T^i = 1 &= z_i^k + \sum_{j=1}^{J} c_{ij} \end{aligned}$$

That is, Y^k represents the data we would hope to observe in list experiments satisfying Imai's three basic identification assumptions with no measurement error but under different structures of the control lists and different population frequencies for the sensitive item.

C "Designed" Lists with $N = 2000$

In this section I report results from Monte Carlo simulations identical to the "designed lists" described in the main text but doubling the "sample size" to $N = 2000$.

As explained in Section 4.1 of the manuscript, I do not expect doubling the sample size to materially alter the findings reported in the main paper for two reasons. First, doubling the sample size does not solve the fundamental conceptual tension between the no liars assumption and the rationale and design of a list experiment. Second, small samples in the extrema are causing computational difficulties and instability. We expect the number of observations in these cells to grow only linearly in N. Doubling the values in Table 3 of the main text, especially under the designed list, would not be of much help. Monte Carlo results largely confirm this.

I present results for the $J = 4$, low prevalence condition here because becasue it represents the most extreme, challenging case for the ICT-MLE that highlights the paper's fundamental contribution. Results here differ from those in the main text in only minor ways that are to be expected with a small increase in the number of observations in the extremes.

C.1 Convergence, stability, and bias in $\hat{b}_1^L$

Table 3 displays the analogue to Figure 2 in the main text. Doubling the sample size increases the number of observations in $\mathcal{J}(J+1)$ about twofold. This has the expected effect of reducing the rate at which the ICT-MLE fails to run, consistent with Figure 2 of the main text.

Table 3: Computational stability and the number of observations in $\mathcal{J}(J+1)$ for the $J=4$, low prevalence condition with $N=2000$.

Error	Mean obs$\in \mathcal{J}(1, J+1)$	% runs with $\mathcal{J}(1, J+1) = \emptyset$	% runs crashed
none	2.2	10.6	1.4
3%	32.1	0.0	0.0

Notwithstanding improved algorithm convergence (i.e., failure to exit with an error), we still observe instability in the actual ICT-MLE regression parameter estimates under the no error condition. Table 4 displays the distribution of estimates for $\hat{b}_1^L$, both with and without error. Recall that the regression parameters are $b_0^L = 0$ and $b_1^L = -4$. The population proportion for the sensitive attribute is $\approx 12\%$. This table is analogous to Table 4 and Figure 4 in the main text. We see that even when the ICT-MLE does run the very small numbers of observations in $\mathcal{J}(1, J+1)$ in the no-error condition make the estimator unstable. Inducing top-biased error stabilizes the estimation but results in upwardly biased $\hat{b}_1^L$. Doubling the sample size had little effect on this result compared to the main text.

Table 4: The distribution of $\hat{b}_1^L$ for the $J=4$, low prevalence condition with $N=2000$. The true parameter value is $b_1^L = -4$.

	no error	3% error
1st Qu.	-12.17	-2.19
Median	-7.75	-1.60
Mean	-204.60	-1.61
3rd Qu.	-5.03	-1.03

C.2 Bias and variance in $\hat{\pi}_{Z^*}^L$

Table 1 displays the distribution of estimates for $\hat{\pi}_{Z^*}^L$ for both DiM and ICT-MLE with and without error. This table mirrors the results presented in Figure 6 of the main text almost exactly.

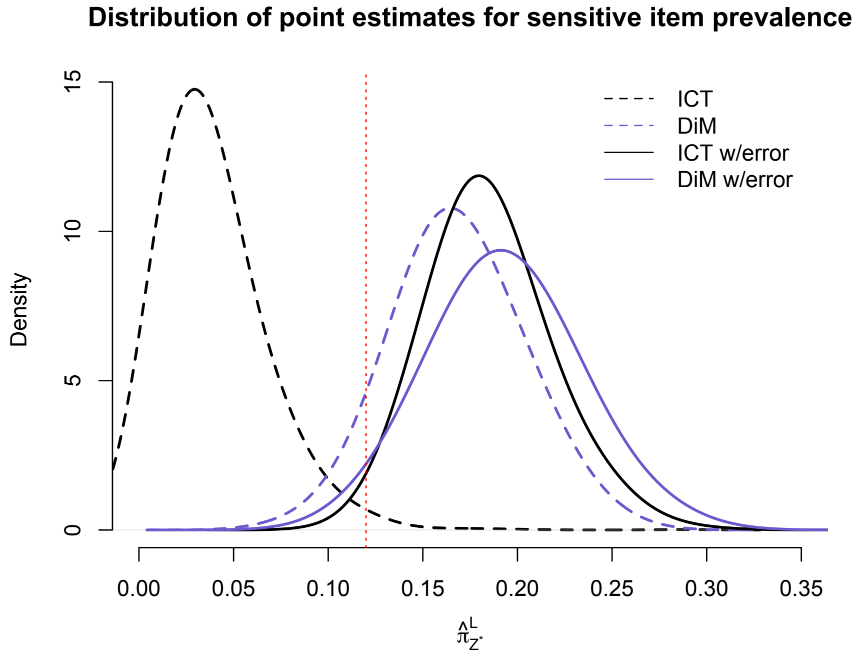

Figure 1: Distribution of point estimates for $\hat{\pi}^L_{Z^*}$ for the DiM and ICT-ML estimators with and without 3% top biased error.

D Further details on the AMJ list experiments

D.1 YouGov Interface for the AMJ Experiments

Figure 2 displays the question (and the YouGov user interface) for the treatment group.

D.2 AMJ calibration lists

References

Ahlquist, John S., Kenneth R. Mayer and Simon Jackman. 2014. "Alien Abduction and Voter Impersonation in the 2012 US General Election: evidence from a survey list experiment." *Election Law Journal* 13(4):460–75.

Blair, Graeme and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20:47–77.

Corstange, Daniel. 2009. "Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT." *Political Analysis* 17:45–63.

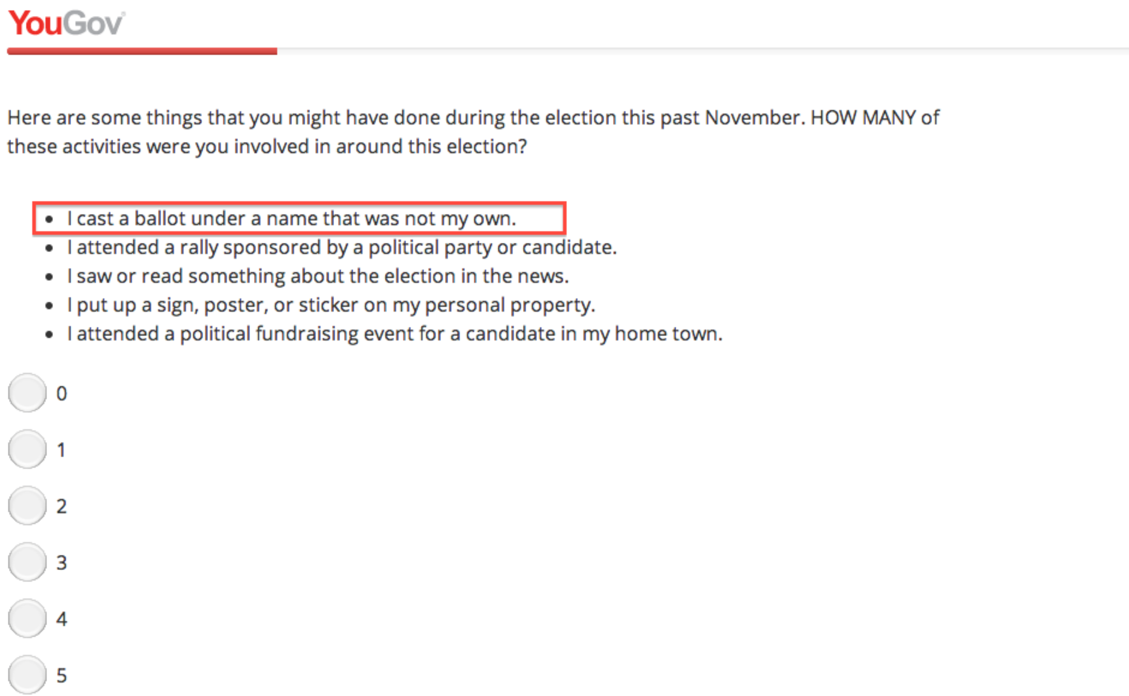

Figure 2: An example of the user interface facing respondents to the YouGov survey employed in Ahlquist, Mayer and Jackman (2014). The sensitive item is highlighted here; actual survey respondents would not see the red box. Respondents in the baseline condition would see only four items with the sensitive item omitted. The ordering of items in the list was randomized.

Table 5: Impossible event list experiment (September 2013 wave)

Prompt:	"Here are some things that may have happened to you during the past twelve months. HOW MANY of these events happened to you?"
1	"I was asked to serve on a jury"
2	"I was called by a telemarketer"
3	"I was audited by the IRS (Internal Revenue Service)"[2]
4	"An airline canceled my flight reservation"
Treatment	"I was abducted by extraterrestrials (aliens from another planet)."

Glynn, Adam. 2013. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77:159–72.

Table 6: Common illegal/undesirable behavior list experiment (September 2013 wave)

Prompt:	"Here are some things that you might have done during the past 30 days. HOW MANY did you do?"
1	"I traveled to a foreign country"
2	"I flossed my teeth"
3	"I littered in a public place"
4	"I celebrated my birthday"
Treatment	"I read or wrote a text (SMS) message while driving"