PA

Relaxing the No Liars Assumption in List Experiment Analyses

Yimeng Li

Graduate Student, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91106, USA. Email: yimeng.li@caltech.edu

Abstract

The analysis of list experiments depends on two assumptions, known as "no design effect" and "no liars". The no liars assumption is strong and may fail in many list experiments. I relax the no liars assumption in this paper, and develop a method to provide bounds for the prevalence of sensitive behaviors or attitudes under a weaker behavioral assumption about respondents' truthfulness toward the sensitive item. I apply the method to a list experiment on the anti-immigration attitudes of California residents and on a broad set of existing list experiment datasets. The prevalence of different items and the correlation structure among items on the list jointly determine the width of the bound estimates. In particular, the bounds tend to be narrower when the list consists of items of the same category, such as multiple groups or organizations, different corporate activities, and various considerations for politician decision-making. My paper illustrates when the full power of the no liars assumption is most needed to pin down the prevalence of the sensitive behavior or attitude, and facilitates estimation of the prevalence robust to violations of the no liars assumption for many list experiment applications.

Keywords: list experiments, partial identification, Monte Carlo simulation

1 Introduction

Eliciting truthful answers from respondents on sensitive issues is a difficult problem in surveys. When surveyors ask respondents these questions directly, some respondents may not be willing to answer truthfully or may even refuse to answer. To address this problem, several survey techniques have been developed and have gained popularity among political scientists and sociologists, such as list experiments (item count technique), randomized response techniques, and endorsement experiments. In political science, scholars have conducted and analyzed list experiments on various topics involving sensitive behaviors or attitudes (early works include Kuklinski, Cobb, and Gilens 1997 on racial prejudice; Kane, Craig, and Wald 2004 on religion; Streb *et al.* 2008 on gender; Corstange 2009 on turnout; and Holbrook and Krosnick 2010 on voting rights; see Section 3.2 for a list of recent works).

List experiments are intended to eliminate the direct association between a respondent's answer and his or her attitude toward the sensitive item by including a few nonsensitive items ("control items").[1] By doing so, list experiments give respondents the opportunity to provide truthful answers to the question while not revealing their attitudes toward the sensitive issue. Even though researchers do not gain knowledge about individual respondents' preferences toward the

Political Analysis (2019)
vol. 27:540–555
DOI: 10.1017/pan.2019.7

Published
10 May 2019

Corresponding author
Yimeng Li

Edited by
Jeff Gill

© The Author(s) 2019. Published by Cambridge University Press on behalf of the Society for Political Methodology.

Author's note: I thank Ines Levin for collecting and sharing with me the data I use in Section 3.1. Previous versions of this research were presented as a poster at the 34th Annual Meeting of the Society for Political Methodology (Polmeth 2017), and in a paper session at the 2018 Annual Meeting of the Midwest Political Science Association (MPSA 2018). I thank R. Michael Alvarez, Jonathan N. Katz, Seo-young Silvia Kim, Ines Levin, Lucas Núñez, Alejandro Robinson-Cortés, Robert Sherman, Matthew Shum, and participants at my poster and paper presentations at Polmeth 2017 and MPSA 2018 for discussions and comments. All errors are my own. Replication data are available in Li (2018) and an R function to implement the proposed bounds is available at https://github.com/lymolympic/list_relaxed_liars.

[1] I describe the structure of list experiments in detail in Section 2.1. In a nutshell, respondents in the control group see a list of control items, and respondents in the treatment group see a list of the sensitive item and control items. Respondents report how many items they would respond to in the affirmative.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

sensitive issue, researchers can obtain the proportion of the population whose truthful responses are affirmative to the sensitive issue under some important assumptions. In their research on list experiments, Imai (2011) and Blair and Imai (2012) formulate the following two assumptions: (1) the inclusion of the sensitive item has no effect on respondents' latent answers to the control items; (2) respondents give truthful latent answers to the sensitive item.[2] The analysis of list experiments depends crucially on these two assumptions, with the first assumption called "no design effect" and the latter referred to as "no liars". The standard estimator is the difference between the average response from treated respondents and that from control respondents. The estimator is simple, intuitive, and unbiased under the assumptions of no design effect and no liars.

The no liars assumption is strong and may fail in many list experiments. In particular, a treated respondent who favors all or a large number of control items and the sensitive item may be reluctant to answer truthfully, as it reveals with certainty or a high probability that the respondent favors the sensitive item. In a recent study, Rosenfeld, Imai, and Shapiro (2016) find that their list experiment underestimates high incidence behavior in the context of a telephone interview. One possible explanation for this observation is that some respondents give untruthful answers for the sensitive item. Statistical tests are useful to detect certain forms of violations, but they usually look at the no design effect assumption, may lack power, or only examine a small subset of respondents. In this paper, I relax the no liars assumption, and in order to do so, maintain the no design effect assumption and focus on behaviors or attitudes with one-sided sensitivity.

I first illustrate which quantities concerning the proportions of different types of respondents are identifiable in the absence of the no liars assumption. In particular, the proportions of truth-telling respondents with the sensitive behaviors or attitudes are identifiable, but liars are observationally equivalent to respondents who do not favor the sensitive item, conditional on the same number of control items answered affirmatively. I proceed to develop a method to estimate bounds for the prevalence of sensitive behaviors or attitudes that does not depend on the no liars assumption. To relax the no liars assumption, I introduce parameters that capture the proportion of liars conditional on the number of control items answered affirmatively. Under a mild condition that the respondents who are supposed to answer affirmatively to all control items as well as the sensitive item have the strongest incentive to lie, I can derive bounds for the level of support for the sensitive item. In particular, the lower bound is the standard difference-in-means estimates. For the upper bound, I first calculate the maximal ratio of liars versus truth tellers among respondents answering affirmatively to all control items, and use this ratio to bound the proportion of liars whose answer is affirmative to fewer than all control items. To permit inference, I construct the confidence set using techniques developed for partially identified models (Imbens and Manski 2004; Stoye 2009).

I then apply my method to a list experiment on the anti-immigration attitudes of California residents and on a broad set of existing list experiment datasets in the literature. In my example on anti-immigration attitudes, I show that I can reach substantive conclusions without having to put full faith on the no liars assumption. I discuss when the bounds under my relaxed liars assumption are likely to be narrow. In particular, my simulated list experiments suggest that positive correlations between the sensitive item and the control items imply narrower bounds. Since such positive correlations are likely to be present when items on a list are of the same category, this result is consistent with my finding of narrower bounds in list experiments conducted in the literature with this feature. The contribution of my paper is to illustrate when the full power of the no liars assumption is most needed to pin down the prevalence of the sensitive behavior or attitude, and to facilitate analysis of list experiments robust to violations of the no liars assumption.

2 Under the first assumption (no design effect), a respondent's answer is the sum of their latent answers to the control items and the sensitive item.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

The structure of the rest of the paper is as follows. I formally describe list experiments and introduce my method in Section 2. In Section 3, I illustrate my method via an example and apply my method to a broad set of list experiment datasets in published studies as well as some simulated list experiments. I discuss a stronger assumption and potential challenges in Section 4 and then conclude.

2 Method

2.1 Setup

Consider a list experiment with N respondents, randomly sampled from the population of interest. Researchers randomly assign the respondents to a control or a treatment condition. The number of respondents under control and treatment need not be the same. Researchers present to the control respondents a list of J control items and ask them how many they would respond to in the affirmative. For treated respondents, researchers ask an otherwise identical question except for the addition of a sensitive item, namely among a list of J control items and a sensitive item how many they would respond affirmatively to.

The canonical difference-in-means estimator is the difference between the average of respondents' answers under treatment and control. Imai (2011) and Blair and Imai (2012) formulate the following two identification assumptions that facilitate the analysis of list experiments:

> Assumption 1 (No design effect). The inclusion of the sensitive item has no effect on respondents' answers to control items.

and

> Assumption 2 (No liars). Respondents give truthful answers for the sensitive item.

Under Assumption 1, the difference in means is simple and intuitive, and it gives an unbiased estimator of the population proportion of those who give an affirmative answer to the sensitive item. Under Assumption 2, it is also the population proportion of those who favor the sensitive item in their underlying preferences.

I make three remarks here. First, following Imai (2011) and Blair and Imai (2012), researchers do not assume respondents give truthful answers to the control items, only that they answer in a consistent way under control and treatment. Second, I make a distinction between a respondent giving an affirmative answer to the sensitive question and the respondent favoring the sensitive item in their underlying preference. This distinction will be important throughout the paper. Finally, in the main text, I describe my method for the case where an affirmative response to the sensitive item is considered sensitive (such as have done, support, or agree with the sensitive item). The case where a negative response is sensitive is analogous and described in Section A.5 in the supplementary materials.

An important implication of no design effect and no liars is that researchers can recover—from the distribution of responses under control and treatment—the joint distribution of respondents' preferences toward the sensitive item and the number of control items answered affirmatively (Glynn 2010). In particular, denote by p_k the proportion of the population who respond affirmatively to k control items and do not favor the sensitive item, and by p_{kT} the proportion responding affirmatively to k control items and favoring the sensitive item (shown in Table 1). Then $(p_k, p_{kT})_{k=0}^{J}$ are identified under Assumptions 1 and 2.[3]

3 To see this, start with p_{JT}. Since under Assumptions 1 and 2 respondents who answer $J+1$ under treatment consist exclusively of those who answer affirmatively to k control items and favor the sensitive item, one can estimate p_{JT} by the proportion giving the maximal answer under treatment. Now respondents who answer J under control consist of those who answer affirmatively to k control items, regardless of their preferences to the sensitive item. So one can estimate p_J by subtracting the estimated p_{JT} from the proportion giving the maximal answer under control. One can continue in this fashion and obtain all $(p_k, p_{kT})_{k=0}^{J}$.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

Table 1. Distribution of different types of respondents ($J = 4$).

	Sensitive item	
Number of control items	No	Yes
0	p_0	p_{0T}
1	p_1	p_{1T}
2	p_2	p_{2T}
3	p_3	p_{3T}
4	p_4	p_{4T}

2.2 No Liars Assumption

Ideally researchers would like all respondents to answer the list experiment question in a consistent and truthful way, satisfying the no design effect assumption and the no liars assumption. However, the no liars assumption is strong and may fail in many circumstances. For a treated respondent whose answer to all control items is affirmative and who favors the sensitive item, truth telling reveals with certainty that the respondent's answer toward the sensitive item is affirmative, often referred to as ceiling effects in the literature (Kuklinski, Cobb, and Gilens 1997; Blair and Imai 2012; Glynn 2013).[4] And for a treated respondent whose answer to a large number of control items is affirmative and who favors the sensitive item, intuitively, truth telling reveals with a high probability that the respondent's answer toward the sensitive item is affirmative. Respondents who feel strongly about the social norm and fear their preferences for the sensitive item might be revealed thus have an incentive to lie about the sensitive item under list experiments, which leads to a violation of the no liars assumption. Moreover, failing to understand the protective nature of the list experiment design, actively managing perceptions about oneself (Köszegi 2006) and being too embarrassed to admit also lead respondents to form untruthful latent response to the sensitive item.

Statistical tests are useful to detect certain violations of the assumptions, but there are limitations to these tests. Blair and Imai (2012) propose to test the positivity of the proportions of different types of respondents, $(p_k, p_{kT})_{k=0}^{J}$. However, they note that "the violation of Assumption 2 [no liars] alone does not lead to negative proportions of these types" and "researchers may fail to reject the null hypothesis due to a lack of statistical power" (Blair and Imai 2012, pp. 64–65). In other words, the utility of the test is primarily detecting failures of the no design effect assumption and lack of statistical power can be a concern. Follow-up work by Aronow *et al.* (2015) proposes a complementary test when a direct question to the same respondents is also available.[5] Their test, however, can only detect violations of list experiment assumptions for respondents who reveal preferences for the sensitive item under the direct question.

In a recent validation study, Rosenfeld, Imai, and Shapiro (2016) find their list experiment underestimates sensitive votes against a referendum. One possible explanation for this observation is that some respondents give untruthful answers for the sensitive item.

In this paper, I will maintain the no design effect assumption and relax the no liars assumption. It will become clear that the no design effect assumption has important implications for identifying different types of respondents on its own and my analysis depends critically on it. For further discussion of this assumption, see Section A.6 in the supplementary materials.

2.3 Identification in the Absence of no Liars Assumption

In the absence of the no liars assumption, the decomposition of respondent types in Table 1 is no longer valid. Instead, I denote the population fraction of respondents who answer affirmatively

4 Similarly, for a treated respondent whose answers to all items—control and sensitive—are negative, truth telling is also fully revealing, known as floor effects.

5 Direct questions also set up useful benchmarks for assessing how well list experiments reduce social desirability bias (Blair and Imai 2012; Kramon and Weghorst 2012).

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

Table 2. Distribution of different types of respondents ($J = 4$).

	Sensitive item		
Number of control items	Nonsupporters	Liars	Truth tellers
0	p_{0N}	p_{0L}	p_{0T}
1	p_{1N}	p_{1L}	p_{1T}
2	p_{2N}	p_{2L}	p_{2T}
3	p_{3N}	p_{3L}	p_{3T}
4	p_{4N}	p_{4L}	p_{4T}

Table 3. Relationship between latent attitudes and observable responses ($J = 4$).

$$
\begin{aligned}
c_0 &= p_{0N} + p_{0L} + p_{0T} & t_0 &= p_{0N} + p_{0L} \\
c_1 &= p_{1N} + p_{1L} + p_{1T} & t_1 &= p_{1N} + p_{1L} + p_{0T} \\
c_2 &= p_{2N} + p_{2L} + p_{2T} & t_2 &= p_{2N} + p_{2L} + p_{1T} \\
c_3 &= p_{3N} + p_{3L} + p_{3T} & t_3 &= p_{3N} + p_{3L} + p_{2T} \\
c_4 &= p_{4N} + p_{4L} + p_{4T} & t_4 &= p_{4N} + p_{4L} + p_{3T} \\
& & t_5 &= p_{4T}
\end{aligned}
$$

to k control items and (1) do *not* favor the sensitive item by p_{kN}; (2) favor the sensitive item, but would *not* give the truthful answer for it by p_{kL}; (3) favor the sensitive item, and give the truthful answer for it by p_{kT}. In particular, the proportion p_k of respondents who answer in the affirmative to k control items now consists of both the proportion p_{kN} of nonsupporters and the proportion p_{kL} of liars. Assuming only one side of the sensitive item is potentially socially undesirable, i.e.,

> *One-sided sensitivity assumption*: Respondents who do not favor the sensitive item do not lie and falsely give an affirmative answer.

I can present the population distribution of different types of respondents in Table 2. In this terminology, the no liars assumption states that $p_{kL} = 0$ for all k.

Researchers do not observe a respondent's latent answer for the control items or the latent attitude toward the sensitive item (except when the respondent gives an answer of 0 or J). What researchers observe is a response from 0 to J for each respondent under the control condition, and a response from 0 to $J + 1$ for each respondent under the treatment condition. Let c_k be the population proportion of respondents whose answer is k under the control condition and t_k be the analogous proportion under the treatment condition, with $\sum_{k=0}^{J} c_k = \sum_{k=0}^{J+1} t_k = 1$.

To map a respondent's latent response toward the control items and the latent attitude toward the sensitive item to his or her answer under the control condition, take a respondent who answers one item as an example. This respondent must answer affirmatively to exactly one control item, but I do not have any information about his or her attitude toward the sensitive item. The respondent may or may not favor the sensitive item, and in the latter case might or might not give a truthful answer for it if he or she were assigned to the treatment condition instead. Therefore, I have $c_1 = p_{1N} + p_{1L} + p_{1T}$. In the same fashion, I can express each c_k in terms of the p-terms, as shown in the left column in Table 3.

For a treated respondent, I can establish the mapping in a similar way. For example, to see $t_1 = p_{1N} + p_{1L} + p_{0T}$, notice that a response of one item under treatment comes from a respondent who (1) answers affirmatively to exactly one control item and does not favor the sensitive item, or (2) answers affirmatively to exactly one control item and favors the sensitive item, but does not give a truthful answer for it, or (3) answers affirmatively to no control items, but favors the sensitive item and is truthful. The expressions for all t_k in terms of the p-terms are shown in the right column in Table 3.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

In Table 3, there are more unknown parameters than equations. Hence, the system of equations is underidentified and I cannot identify the distribution of all latent types, $(p_{kL}, p_{kN}, p_{kT})_{k=0}^{J}$. A closer inspection reveals the reason why this system is underidentified. A lying respondent who actually favors the sensitive item is observationally equivalent to a respondent who does not favor the sensitive item at all, conditional on the same number of control items answered affirmatively. As a result, only the sum of the proportions of these two types is identified. In particular, regarding $p_{kN}+p_{kL}$ as one parameter instead of two parameters p_{kN} and p_{kL}, the system is exactly identified and I can identify $p_{kN} + p_{kL}$ and p_{kT} for each $k = 1, \ldots, J$.

2.4 Relaxed Liars Assumption, Bounds and Inference

The quantity of interest is the population proportion of respondents who favor the sensitive item, $\sum_{k=0}^{J} p_{kL}+p_{kT}$. I have already shown that the proportion of truth tellers who respond affirmatively to k control items, p_{kT}, is identified for each k, the remaining work concerns the fraction of respondents who favor the sensitive item but do not give truthful answers for it, $\sum_{k=0}^{J} p_{kL}$. Since $p_{kN} + p_{kL}$ is identified for each $k = 1, \ldots, J$, by setting all p_{kL}'s to be zero (no liars), I obtain the standard difference-in-means estimator as a lower bound of the level of support for the sensitive item. At the other extreme, if I set all p_{kN}'s to be zero (no nonsupporters), I obtain a trivial upper bound of 1.

With no behavioral assumption about the pattern of lying, I do not have an informative upper bound on the prevalence of the sensitive item. Toward a nontrivial upper bound, I propose the following assumption concerning the incentive to lie:

Relaxed liars assumption:

$$\frac{p_{kL}}{p_{kL} + p_{kT}} \leq \frac{p_{JL}}{p_{JL} + p_{JT}}, \quad \forall k = 0, \ldots, J-1, \tag{1}$$

where $p_{kL}/(p_{kL} + p_{kT})$ is the ratio of liars conditional on responding affirmatively to k control items and favoring the sensitive item. This assumption says that among all respondents who favor the sensitive item, the ones who respond affirmatively to all control items have the strongest incentive to lie. It relaxes the assumption of no liars, which amounts to $p_{kL}/(p_{kL} + p_{kT}) \equiv 0$ for all $k = 0, \ldots, J$. The relaxed liars assumption is simple and intuitive, and it captures the intuition behind the ceiling effect that truth telling reveals with certainty a respondent's preference for the sensitive item when he or she favors all items. Also note that this assumption does not impose any relationship between lying incentives for respondents favoring fewer than all control items.

My relaxed liars assumption gives an upper bound for the number of liars among respondents who favor fewer than all control items. However, I need an estimate or upper bound on p_{JL} to make this assumption implementable. Since $p_{JN} + p_{JL}$ is identified, it can serve as an upper bound. This bound will be crude unless p_{JN} is relatively small. As presented in Section 3, in some applications, the crude bound can nevertheless be informative. In other list experiments, the bound can be wide, in which case the no liars assumption is important to nail down the prevalence of the sensitive item. Nonzero lower bound on p_{kN}, $k = 0, \ldots, J$, would improve upon the crude bounds, but would also necessarily involve additional information and/or assumptions. Formally, an upper/lower bound for the level of support for the sensitive item is given by the solution to the following linear system:

$$\begin{aligned}
&\max_{(p_{kL},p_{kN},p_{kT})_{k=0}^{J}} \Big| \min_{(p_{kL},p_{kN},p_{kT})_{k=0}^{J}} \sum_{k=0}^{J} p_{kL} + p_{kT} \\
&\text{s.t. } c_0 = p_{0N} + p_{0L} + p_{0T}, \ldots, c_J = p_{JN} + p_{JL} + p_{JT} \\
&\quad t_0 = p_{0N} + p_{0L}, \ldots, t_J = p_{JN} + p_{JL} + p_{J-1,T},\ t_{J+1} = p_{JT} \\
&\quad p_{kL}/(p_{kL} + p_{kT}) \leq (p_{JN} + p_{JL})/(p_{JN} + p_{JL} + p_{JT}), \quad \forall k = 0, \ldots, J-1.
\end{aligned} \tag{2}$$

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

I derive the bounds in steps in Section A.1 in the supplementary materials, and the lower and upper bounds are given by

$$\left[\sum_{k=0}^{J} p_{kT}, \sum_{k=0}^{J} p_{kT} + \sum_{k=0}^{J} \min\left\{\frac{\lambda}{1-\lambda} p_{kT}, p_{kN} + p_{kL}\right\}\right] \tag{3}$$

where $p_{kT} = \sum_{i=k+1}^{J+1} t_i - \sum_{j=k+1}^{J} c_j$, $p_{kN} + p_{kL} = \sum_{j=k}^{J} c_j - \sum_{i=k+1}^{J+1} t_i$, and the maximal liar ratio $\lambda \equiv (p_{JN}+p_{JL})/(p_{JN}+p_{JL}+p_{JT}) = (c_J - t_{J+1})/c_J$. It follows that the bounds are nontrivial if and only if $\frac{\lambda}{1-\lambda} p_{kT} \leq p_{kN} + p_{kL}$, or equivalently $p_{kT}/(p_{kN}+p_{kL}) \leq p_{JT}/(p_{JN}+p_{JL})$ for some k. In words, the condition holds if the proportion of truth-telling supporters is higher for respondents answering affirmatively to all J control items than $k < J$ control items. A higher prevalence of the sensitive item will increase $\{p_{kT}\}_{k=0}^{J}$ but decrease λ, so the overall effect is ambiguous. Factors holding the prevalence fixed while decreasing the maximal liar ratio will narrow the bounds. I conduct detailed investigations in Section 3.

After identifying the lower and upper bounds, which I denote by $\hat{p}_l$ and $\hat{p}_u$, I can construct the $(1-\alpha)$-confidence intervals $\text{CI}(\hat{p}_l)$ and $\text{CI}(\hat{p}_u)$ via bootstrap. The confidence set formed by the lower endpoint of $\text{CI}(\hat{p}_l)$ and the upper endpoint of $\text{CI}(\hat{p}_u)$ covers the identified set with probability $(1-\alpha)$, but will be slightly conservative in terms of covering the true prevalence of the sensitive item. Alternatively, I can construct the confidence set with $(1-\alpha)$-coverage following Imbens and Manski (2004) and Stoye (2009) as

$$\text{CI}_\alpha = \left[\hat{p}_l - c_\alpha \frac{\hat{\sigma}_l}{\sqrt{N}}, \hat{p}_u + c_\alpha \frac{\hat{\sigma}_u}{\sqrt{N}}\right], \tag{4}$$

where N is the total sample size, $\hat{\sigma}_l$ ($\hat{\sigma}_u$) is the standard error for $\hat{p}_l$ ($\hat{p}_u$), and c_α solves

$$\Phi\left(c_\alpha + \frac{\sqrt{N}}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}}(\hat{p}_u - \hat{p}_l)\right) - \Phi(-c_\alpha) = 1 - \alpha, \tag{5}$$

where Φ is the standard normal cumulative distribution function.[6,7] For a 95% confidence set, c_α will be $\Phi^{-1}(0.975) \approx 1.96$ if $\hat{p}_u - \hat{p}_l = 0$ and will approach $\Phi^{-1}(0.975) \approx 1.64$ as $\hat{p}_u - \hat{p}_l$ grows large relative to sampling error.

2.5 Extensions

2.5.1 *Auxiliary Information to Sharpen Bounds*

Further information can help us bound p_{kN} and improve upon the crude upper bound. Auxiliary direct questions can be helpful to sharpen the bound. For example, researchers interested in learning the prevalence of anti-immigration attitudes can ask a direct question on support for a pathway to citizenship with possible choices yes, no, and ambivalent answers (not sure, rather not say, or do not know). The proportion of yes answers—indicating pro-immigration attitudes—among those who favor k control items is a conservative estimate for p_{kN}, under the assumption that the proportion of liars who answer yes to this direct question is smaller than that of nonsupporters who give negative or ambivalent answers.

6 Notice that the upper bound is continuous but not differentiable at points satisfying $p_{kT} \cdot \lambda/(1-\lambda) = p_{kN} + p_{kL}$. Outside this set of measure zero, the upper bound is differentiable and the delta method gives asymptotic normality pointwise. Moreover, outside neighborhoods of these nondifferentiability points (that can be taken to have arbitrarily small measure), standard argument gives uniform convergence to normal distribution asymptotically.

7 In principle, one can calculate the asymptotic variance of $\hat{p}_l$ and $\hat{p}_u$, and obtain $\hat{\sigma}_l$ and $\hat{\sigma}_u$ by taking sample analogues. I avoid the cumbersome derivation by bootstrapping.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

2.5.2 *Too Few Respondents on the Boundary*

To reduce the number of respondents choosing zero or all items, some applied researchers include an item of very high prevalence and an item of very low prevalence. Glynn (2013) instead suggests including negatively correlated items in favor of using high- or low-prevalence items. Both practices may result in very few respondents favoring all control items in some list experiments. In this case, the estimation of maximal liar ratio $(p_{JN} + p_{JL})/(p_{JN} + p_{JL} + p_{JT})$ in (2) relies on a small number of observations.

If answering affirmatively to all control items is vanishingly unlikely, then answering affirmatively to all but one control items is effectively an extreme response. So instead of estimating the maximal liar ratio as $(p_{JN} + p_{JL})/(p_{JN} + p_{JL} + p_{JT})$ in (2), I collapse the case of all control items and all but one control items and replace it with $(p_{J-1,N} + p_{J-1,L} + p_{JN} + p_{JL})/(p_{J-1,N} + p_{J-1,L} + p_{J-1,T} + p_{JN} + p_{JL} + p_{JT})$, a quantity that can be more precisely estimated.[8] In words, I now require respondents favoring at least all but one control items to have stronger incentive to lie than those favoring fewer items.

3 Applications

3.1 Illustrative Example

For illustration, consider a list experiment Alvarez *et al.* (2019) included in an online survey of 2,722 California adults conducted in 2014 using a sample recruited by Qualtrics. I excluded inattentive respondents who fail one of three trap questions included in the survey, and this leaves me with a total of 1,750 respondents.[9] The experiment was designed to measure the support for two state chapters of national anti-immigrant organizations among California residents. To preserve the anonymity of these organizations, I refer to them here as organization X and organization Y.

Respondents saw a control list or a treatment list with the names and short descriptions of different groups and organizations as shown in Table A.2, and were instructed to specify "how many of these groups and organizations you broadly support".

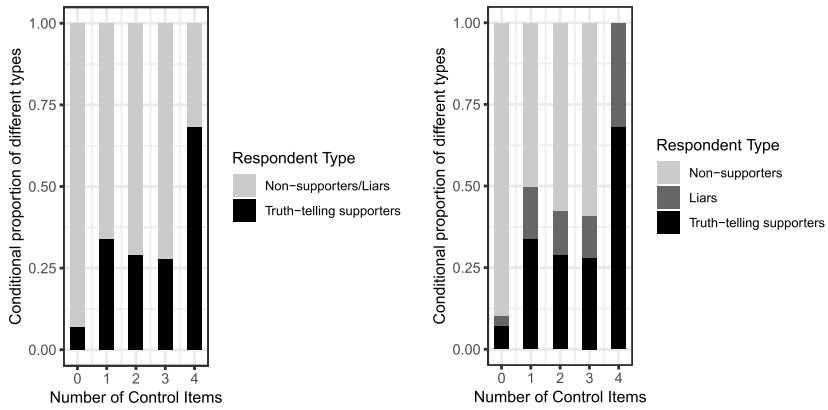
The proportions of control and treated respondents giving each answer are shown in Table A.3. Under the no design effect assumption, I can identify the proportions of truth-telling supporters and (truth-telling) nonsupporters/liars conditional on the number of control items answered affirmatively. Figure 1 shows these proportions for organization X on the left. If I further impose the relaxed liars assumption and allow a maximal number of liars, I can obtain the proportion of truth-telling supporters, the maximal proportion of liars, and the minimal proportion of nonsupporters, all conditional on the number of control items answered affirmatively. These proportions are shown in Figure 1 on the right.

Doing the calculations, the difference-in-means estimates are 0.36 for Organization X and 0.22 for Organization Y. 95% confident intervals for the point estimates are $[0.19, 0.53]$ and $[0.05, 0.39]$, respectively. From my argument in Section 2, the difference-in-means estimate may understate the level of support for these organizations by neglecting respondents who support them but did not give truthful answers. If I allow the possibility of liars, but assume my relaxed liars assumption and implement my crude bound, I obtain an upper bound 0.53 for the level of support for Organization X, and 0.36 for Organization Y. 95% confidence sets for the interval estimates are $[0.23, 0.80]$ and $[0.10, 0.64]$ for Organization X and Organization Y, respectively.

8 When there are "too few" respondents answering the maximal number of items and to collapse the two cases is a subjective decision to make. For applications in Section 3, most list experiments have a sample size of 1,000–2,000 with a few notable exceptions. I decide to collapse the two cases whenever there are fewer than 50 respondents choosing all items (or zero items for list experiment with negative sensitive responses).

9 Alvarez *et al.* (2019) implemented a double list experiment in fixed list order but for my purposes, I focus on the first list in this paper. I refer interested readers to Alvarez *et al.* (2019) for details on these trap questions and the differences in the response pattern between respondents who passed and failed the trap questions, in terms of their reported political knowledge, political participation, ideology, and responses to the list experiment.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press



Note: The figure presents, for Organization X (the sensitive item), the proportions of truth-telling supporters and non-supporters/liars (left), and the proportions of truth-telling supporters, liars, and non-supporters under the relaxed liars assumption with a maximal number of liars (right), conditional on the number of control items answered affirmatively.

Figure 1. Proportions of different types of respondents for Organization X.

My bound estimates have important implications. To reach the conclusion that at most around a half respondents support Organization X and around a third respondents support Organization Y, I do not need to have full faith on the no liars assumption.

3.2 List Experiments in Published Studies

I now turn attention to a comprehensive set of list experiments in the political science literature, with basic information about these studies summarized in Table A.5.[10]

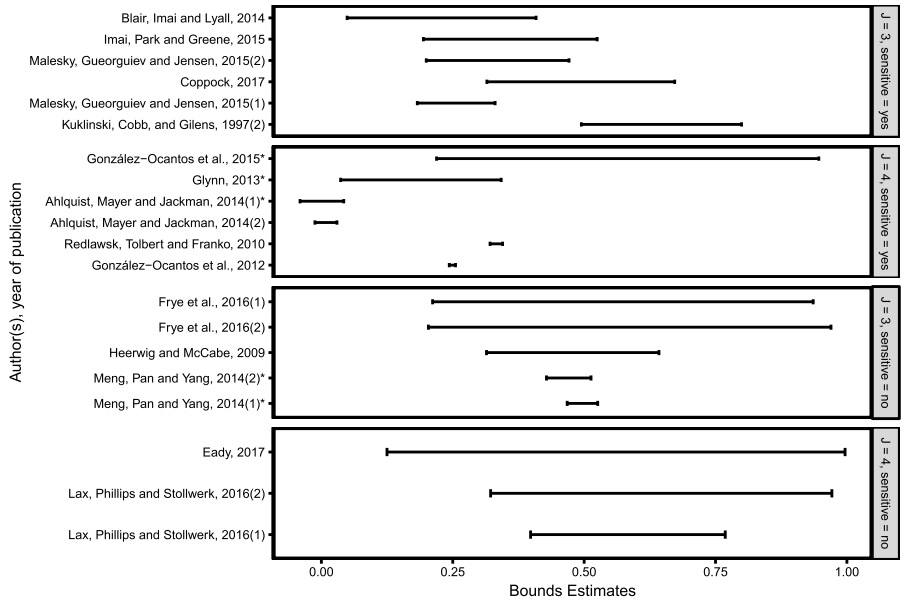
First notice that the list experiment concerning racial prejudice in Kuklinski, Cobb, and Gilens (1997) and the list experiment concerning vote for an anti-abortion referendum in Rosenfeld, Imai, and Shapiro (2016) fail the Blair–Imai test of no design effect. Therefore, I exclude these applications from subsequent analysis. I now apply my method to obtain the range of the prevalence of sensitive behaviors or attitudes possible under the relaxed liars assumption for all other list experiments. Since no extra information for constructing tighter bounds for these studies is available, I compute my crude bounds. Recall that the lower bounds correspond to the difference-in-means estimates under the no liars assumption. In Figure 2, the intervals show the estimated lower and upper bounds.[11] The top two panels and the bottom two panels display studies with affirmative and negative sensitive responses, respectively. Within each panel, the studies are ordered by the maximal proportion of liars permitted by the relaxed liars assumption (with the largest value on the top).

The widths of the intervals vary substantially across studies. Within each panel, the intervals tend to be narrower for studies with a smaller maximal proportion of liars permitted by the relaxed liars assumption. Across panels, the intervals appear to be narrower for list experiments with positive sensitive responses than those with negative sensitive responses, with the exception

10 These studies are: Kuklinski *et al.* (1997), Heerwig and McCabe (2009), Redlawsk *et al.* (2010), González-Ocantos *et al.* (2012), Glynn (2013), Ahlquist *et al.* (2014), Blair *et al.* (2014), Meng *et al.* (2014), González-Ocantos *et al.* (2015), Imai *et al.* (2015), Malesky *et al.* (2015), Frye *et al.* (2016), Lax *et al.* (2016), Coppock (2017), and Eady (2017). I include—to the best my knowledge—all published list experiments conducted in political science with available replication data or the distribution of responses in the paper before 2017, with the exception of Aronow *et al.* (2015), Kiewiet de Jonge (2015) and Coffman, Coffman, and Marzilli Ericson (2017) as they include multiple list experiments ($\geq$8) and I want to put together list experiments from a wide range of surveys and topics.

11 95% confidence sets for these bounds are suppressed to keep the figure uncluttered. As is the case for my illustrative example, the confidence sets are wide in general. The notable exceptions are Malesky, Gueorguiev, and Jensen (2015), Coppock (2017), and Eady (2017) as these list experiments are implemented in larger scale.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press



Note: For list experiments marked with stars, procedures described in Section 2.5.2 are adopted.

Figure 2. Bound estimates for list experiments in published studies.

of Meng, Pan, and Yang (2014) and González-Ocantos, Kiewiet de Jonge, and Nickerson (2015). However, this pattern can be due to difference in other aspects of these list experiments; in fact, there is a symmetry between the case of an affirmative sensitive item and a negative sensitive item.[12] To better understand why some intervals are narrow and some are wide, I take a closer look at these list experiments.

As I showed in Section 2.3 and illustrated in the example, under no design effect, I can identify the proportions of (1) truth tellers with the sensitive behavior or attitude p_{kT}, and (2) the combination of liars p_{kL} and truth tellers without the sensitive behavior or attitude p_{kN}, conditional on the number of control items answered affirmatively (or negatively for list experiments with negative sensitive responses). Figures A.1 and A.2 display these conditional proportions for studies with affirmative and negative sensitive responses, respectively. In other words, the lines in these figures correspond to the line connecting the top of the black bars shown in Figure 1. Since crude bounds use no external information about liars versus nonsupporters for sensitive items, narrower bounds come from: first, increasing proportions of truth tellers with the sensitive behavior/attitude conditional on the number of control items answered affirmatively/negatively, which corresponds to the inequalities in the relaxed liars assumption (1); second, a smaller maximal proportion of liars permitted by the relaxed liars assumption, which corresponds to the right hand side of (1). As shown in Figures A.1 and A.2, the increasing trend is present for many list experiments with affirmative sensitive responses and several with negative sensitive responses. This pattern explains why the intervals in Figure 2 appear to be narrower for these studies. Meanwhile, the widths of the intervals in Figure 2 are governed by the maximal possible proportion of liars in Figures A.1 and A.2, which in turn depends on the prevalence of different control and sensitive items that varies across studies.

One important implication is that the crude bound is more informative when the control items and the sensitive item are of the same type, which may give rise to positive correlation among

12 Simulation results described in the next section also suggest whether the sensitive response is affirmative or negative does not drive the bound widths by itself.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

different items. In a list experiment included in a large-scale administrative survey in Malesky, Gueorguiev, and Jensen (2015), items on the control and treatment lists are different activities that firms engage in to expedite the steps needed to receive their investment license/registration certificate. Another notable example is Meng, Pan, and Yang (2014), who study the factors that Chinese government officials consider when making local policy and expenditure decisions. Control items include considerations such as local administrative expenditures, influence in attracting foreign investment and scope of the migrant population, and the sensitive items are suggestions from residents expressed through formal state institutional channels and those through the Internet. Given the items included in these list experiments, it is not surprising that the crude bounds are informative, as is clear from Figure 2.

3.3 Simulated List Experiments

I now investigate the determinants of the length of the intervals through several simulated list experiments. The benefit of looking at these artificially generated list experiments is that I have full knowledge of the data generating processes, which is not the case for list experiments in published studies. I focus on the differences in the data generating processes and the resulting differences in the interval lengths, and not absolute lengths.

I consider a sensitive item of high prevalence (0.62) and list experiments of length 4 with 2,000 respondents in total, and vary the design of the control-item list and the distribution of liars conditional on the number of control items answered affirmatively.

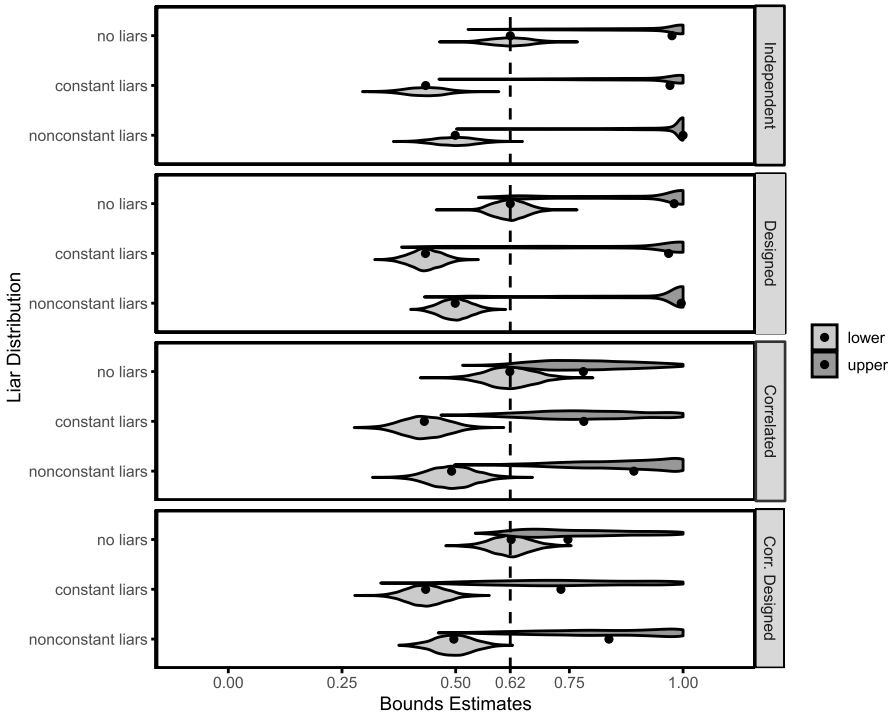
I study four list structures. In an "independent" list, all control items are independent with prevalence $(1/6, 1/2, 2/3, 2/3)$, and independent from the sensitive item. In a "designed" list, there are high- and low-prevalence control items (0.85 and 0.15), as well as two negatively correlated control items, with prevalence 0.5 and correlation -0.6. These parameter choices are borrowed from Ahlquist (2017). I modify the "independent" lists and "designed" lists by adding 0.1 to the correlation of each pair of (control and sensitive) items, and refer to them as "correlated" and "correlated designed" lists.

I look at three different liar distributions. In the case of "no liars", the lower bound—the difference-in-means estimate—consistently estimates the prevalence of the sensitive item. In the case of "constant liars", I assume 30% of the respondents who favor the sensitive item answer the question as if they do not favor it. Finally, in the case of "nonconstant liars", I assume $30\% \times 0.8^k$ liars for respondents who answer affirmatively to all but k control items.

Figure 3 displays the densities and the medians of the crude bound estimates (lower bounds in lighter gray and upper bounds in darker gray) for 2,000 simulated datasets for each of the 12 types of list experiments under consideration. As expected, in the case of "no liars", the lower bounds consistently estimate the prevalence of the sensitive item. In the case of "constant liars" and "nonconstant liars", the lower bounds underestimate the prevalence of the sensitive item by the proportions of liars in the population, whereas the interval estimates cover the prevalence. If the sensitive item is truly independent of control items, then the prevalence of sensitive item conditional on the number of control items answered affirmatively is flat. And as the top half of the figure shows, I have trivial upper bounds most of the time for "independent" and "designed" lists. On the other hand, if there is positive correlation between the sensitive item and control items, then the prevalence of sensitive item conditional on the number of control items answered affirmatively is increasing. And as the bottom half of the figure shows, I have nontrivial upper bounds for "correlated" and "correlated designed" lists.

I present further simulations in the supplementary materials. The widths of the bounds are decreasing in the correlation between list items (Figures A.3 and A.4), consistent with our observation in Section 2.4 that decreasing the maximal liar ratio while keeping the prevalence fixed narrows the bounds. On the other hand, whether the sensitive response is affirmative or

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press



Note: The figure presents the densities of lower bound and upper bound estimates for 2000 simulated datasets for each of the 12 types of list experiments. Each shaded area has measure one. Lower bound densities are shown in lighter grey and upper bound densities are shown in darker grey. Black dots are the median lower bound and upper bound estimates.

Figure 3. Bound estimates for simulated list experiments.

negative has no effect on the bound widths (Figure A.6), suggesting the differences between studies with affirmative and negative sensitive responses in Section 3.2 are due to differences in other aspects of these list experiments. Finally, list experiments with lower-prevalence items tend to produce slightly wider bounds in my simulations (Figure A.5).

3.4 Summary

By construction, the bounds will be nontrivial if and only if the proportion of truth-telling supporters is higher for respondents answering affirmatively to all J control items than $k < J$ control items for some k. The stronger the tendency that the proportion of truth-telling supporters increases in the number of control items answered affirmatively (especially when it approaches the boundary), the more informative the crude bounds are, as applications illustrate.

To put in primitive terms, the simulation exercise shows that positive correlation between the sensitive item and control items overall leads to narrower bounds, which is consistent with my previous observation that the width of the crude bound is narrower when items are of the same type. Absent such positive correlation, I expect uninformative bounds. For a list experiment with auxiliary question asking about the sensitive item directly, in addition to visualization in the form of Figure 1, researchers can also learn about the correlation by examining the correlation between control respondents' answers to list experiment and their answers to the auxiliary question.[13]

13 This requires the propensity to lie about the sensitive item under the auxiliary question to be uncorrelated with the number of control items answered affirmatively under the list experiment for control respondents.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

4 Discussion

4.1 Strengthening the Relaxed Liars Assumption

The relaxed liars assumption is a weaker assumption than the no liars assumption, leaving open the possibility that assumptions lying in between may give a tighter bound while accommodating the potential presence of liars. One such assumption states that the proportion of liars is increasing in the number of control items answered affirmatively:

$$\frac{p_{kL}}{p_{kL}+p_{kT}} \leq \frac{p_{mL}}{p_{mL}+p_{mT}}, \quad \forall 0 \leq k < m \leq J-1. \tag{6}$$

Without further information or assumption about the proportion of nonsupporters, p_{mL}, one can weaken the inequality by replacing $p_{mL}/(p_{mL}+p_{mT})$ with $(p_{mL}+p_{mN})/(p_{mL}+p_{mN}+p_{mT})$. Hence, the improvement of this stronger assumption over the relaxed liars assumption depends on the relationship between $(p_{mL}+p_{mN})/(p_{mL}+p_{mN}+p_{mT})$ and $(p_{JL}+p_{JN})/(p_{JL}+p_{JN}+p_{JT})$. Since $(p_{JL}+p_{JN})/(p_{JL}+p_{JN}+p_{JT})$—one minus the proportion of truth-telling supporters conditional on all control items answered affirmatively—is smaller for many list experiments in Section 3.2, the gain is limited for these applications.

4.2 Nonstrategic Respondent Error

By looking at violations of the no liars assumption, I essentially restrict attention to strategic misreporting. In addition to strategic misreporting, an inattentive respondent engaging in survey satisficing may choose randomly an answer, the first answer or the last answer from all options available, suggested by evidence from placebo list experiments (Holbrook and Krosnick 2010; Ahlquist, Mayer, and Jackman 2014; Kiewiet de Jonge and Nickerson 2014).[14] Alvarez *et al.* (2019) find in two datasets with screener questions, difference-in-means estimates for respondents failing screener questions are significantly different from those for attentive respondents.[15]

Nonstrategic respondent error can lead to violations of no design effect, no liars, and my relaxed liars assumption, result in bias to the difference-in-means estimator and maximum likelihood estimator à la Imai (2011) (Ahlquist 2017), and undermine the validity of the bounds developed in this paper. For online surveys, I recommend that applied researchers use screener questions—instructional manipulation checks (Oppenheimer, Meyvis, and Davidenko 2009; Berinsky, Margolis, and Sances 2014) or conventional instructed-response items (Alvarez *et al.* 2019)—to identify inattentive respondents so as to minimize such biases whenever possible. For face-to-face or telephone surveys (where survey satisficing is arguably less serious), screener questions are infeasible and researchers may have to rely on response time or other indicators to identify inattentive respondents.

4.3 Heterogeneous Lying Patterns

Some subgroups of respondents may be more truthful than others. If a respondent characteristic is correlated with both the number of control items answered affirmatively and lying about the sensitive item, then depending on the lying pattern and the magnitude of such correlation, it could potentially violate the relaxed liars assumption. For example, suppose respondents with less education are less likely to realize when ceiling effects are occurring and are more truthful as a result. Then the relaxed liars assumption will be violated if they are disproportionately

14 A placebo list experiment replaces the sensitive item in a standard list experiment with an item whose prevalence is either known or estimable.

15 In speculation to why such inflation/deflation might happen, Kiewiet de Jonge and Nickerson (2014) mention "[g]iven that the ICT tends to require greater cognitive effort than direct questions (particularly in the case of live-interviewer modes) and some respondents engage in satisficing strategies". Ahlquist, Mayer, and Jackman (2014) find some suspicious respondents display straight-lining patterns.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

more likely to answer affirmatively to all control items, and the magnitude is large enough to offset the strategic incentive underlying the assumption. Researchers can check whether respondents answering affirmatively to all control items are very different in composition from other respondents, and make changes to the control items if possible (for example, after a pilot study). If there is still concern, researchers can apply the method to each demographic group separately (estimation precision will be lower).

5 Conclusion

Standard analysis of list experiments requires a strong assumption on respondents' truthfulness toward sensitive issues, known as no liars. In this paper, I derive bounds for the prevalence of sensitive behaviors or attitudes under a weaker behavioral assumption and apply my method to an example on anti-immigration attitudes and a broad set of list experiments in the literature. The widths of the bounds vary substantially across studies. In some cases, especially when items on the list are of the same category, I can still reach substantive conclusions without putting full faith on the no liars assumption. In other cases, the full power of the no liars assumption is needed to pin down the prevalence of the sensitive behaviors or attitudes. I suggest that applied researchers compute the bounds as a guard against potential existence of liars in list experiments.

List experiments are designed to incentivize more truth telling from respondents than direct questions, but how successfully any particular application achieves it is not clear.[16] In particular, a small incremental support for the sensitive item over direct questions can mean either a small social desirability bias, or an unsuccessful inducement of truth telling. The method proposed in this paper is a first attempt to accommodate the presence of liars. The bounds exploit underlying strategic incentives and information in list experiments, but can be uninformative depending on applications. When there are resources, applied researchers can also employ other strategies such as the randomized response technique and compare estimates to gain confidence.

The method I propose is an addition to the list experiment toolbox and applicable to a large number of list experiments. But post-data-collection methods should not substitute or be prioritized over efforts to design good list experiments in the first place. Applied researchers should carefully construct the list items so that the question looks natural and the sensitive item stands out to a lesser extent, and follow other recommended practices discussed in Blair and Imai (2012) and Glynn (2013). Meanwhile, when the list features a positive correlation between the sensitive item and control items overall, researchers can exploit this opportunity to estimate bounds robust to the presence of liars.

Supplementary material

For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2019.7.

References

Ahlquist, J. S. 2018. "List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators." *Political Analysis* 26(1):34–53.

Ahlquist, J. S., K. R. Mayer, and S. Jackman. 2014. "Alien Abduction and Voter Impersonation in the 2012 U.S. General Election: Evidence from a Survey List Experiment." *Election Law Journal: Rules, Politics, and Policy* 13(4):460–475.

Alvarez, R. M., L. R. Atkeson, I. Levin, and Y. Li. 2019. "Paying Attention to Inattentive Survey Respondents." *Political Analysis* 27(2):145–162.

16 Another issue worth pointing out is that there is potentially a file drawer problem in list experiment research, like many other areas of scientific research.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

Aronow, P. M., A. Coppock, F. W. Crawford, and D. P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3(1):43–66.
Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.
Blair, G., and K. Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1):47–77.
Blair, G., K. Imai, and J. Lyall. 2014. "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science* 58(4):1043–1063.
Coffman, K. B., L. C. Coffman, and K. M. Marzilli Ericson. 2017. "The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated." *Management Science* 63(10):3168–3186.
Coppock, A. 2017. "Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-Representative List Experiment." *Statistics, Politics and Policy* 8(1):29–40.
Corstange, D. 2009. "Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT." *Political Analysis* 17(1):45–63.
Eady, G. 2017. "The Statistical Analysis of Misreporting on Sensitive Survey Questions." *Political Analysis* 25(2):241–259.
Frye, T., S. Gehlbach, K. L. Marquardt, and O. J. Reuter. 2016. "Is Putin's Popularity Real? *Post-Soviet Affairs* 33(1):1–15.
Glynn, A. N. 2010. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." Unpublished manuscript.
Glynn, A. N. 2013. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77(S1):159–172.
González-Ocantos, E. *et al.* 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.
González-Ocantos, E., C. Kiewiet de Jonge, and D. W. Nickerson. 2015. "Legitimacy Buying: The Dynamics of Clientelism in the Face of Legitimacy Challenges." *Comparative Political Studies* 48(9):1127–1158.
Heerwig, J. A., and B. J. McCabe. 2009. "Education and Social Desirability Bias: The Case of a Black Presidential Candidate." *Social Science Quarterly* 90(3):674–686.
Holbrook, A. L., and J. A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1):37–67.
Imai, K. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494):407–416.
Imai, K., B. Park, and K. F. Greene. 2015. "Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models." *Political Analysis* 23(2):180–196.
Imbens, G. W., and C. F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72(6):1845–1857.
Kane, J. G., S. C. Craig, and K. D. Wald. 2004. "Religion and Presidential Politics in Florida: A List Experiment." *Social Science Quarterly* 85(2):281–293.
Kiewiet de Jonge, C. P. 2015. "Who Lies About Electoral Gifts? Experimental Evidence from Latin America." *Public Opinion Quarterly* 79(3):710–739.
Kiewiet de Jonge, C. P., and D. W. Nickerson. 2014. "Artificial Inflation or Deflation? Assessing the Item Count Technique in Comparative Surveys." *Political Behavior* 36(3):659–682.
Köszegi, B. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4(4):673–707.
Kramon, E., and K. R. Weghorst. 2012. "Measuring Sensitive Attitudes in Developing Countries: Lessons from Implementing the List Experiment." *Newsletter of the APSA Experimental Section* 3(2):14–24.
Kuklinski, J. H., M. D. Cobb, and M. Gilens. 1997. "Racial Attitudes and the 'New South'." *The Journal of Politics* 59(2):323–349.
Lax, J. R., J. H. Phillips, and A. F. Stollwerk. 2016. "Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment." *Public Opinion Quarterly* 80(2):510–533.
Li, Y. 2018. "Replication Data for: Relaxing the No Liars Assumption in List Experiment Analyses." https://doi.org/10.7910/DVN/DFRSKZ, Harvard Dataverse, V1.
Malesky, E. J., D. D. Gueorguiev, and N. M. Jensen. 2015. "Monopoly Money: Foreign Investment and Bribery in Vietnam, a Survey Experiment." *American Journal of Political Science* 59(2):419–439.
Meng, T., J. Pan, and P. Yang. 2014. "Conditional Receptivity to Citizen Participation: Evidence From a Survey Experiment in China." *Comparative Political Studies* 50(4):399–433.
Oppenheimer, D. M., T. Meyvis, and N. Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45(4):867–872.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press

Redlawsk, D. P., C. J. Tolbert, and W. Franko. 2010. "Voters, Emotions, and Race in 2008: Obama as the First Black President." *Political Research Quarterly* 63(4):875–889.

Rosenfeld, B., K. Imai, and J. N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* 60(3):783–802.

Stoye, J. 2009. "More on Confidence Intervals for Partially Identified Parameters." *Econometrica* 77(4):1299–1315.

Streb, M. J., B. Burrell, B. Frederick, and M. A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72(1):76–89.

https://doi.org/10.1017/pan.2019.7 Published online by Cambridge University Press