

论列举实验在敏感问题调查中的应用 ——以非制度化政治参与为验证

王浦劬 季程远

(北京大学 政府管理学院 北京 100871)

摘要: 敏感问题测量是社会科学实证研究的难点, 基于社会意愿偏差, 人们对于敏感问题或者选择不应答, 或者选择非真实应答。随着调查研究技术的进步, 近年来, 社会科学界为此掀起讨论热潮, 列举实验是人们讨论的热门方法之一。列举实验的关键在于避免产生设计效应、天花板效应和地板效应。本文以非制度化政治参与的调查数据分析为基础, 验证显示列举实验可以降低无应答比例, 提高真实应答比例, 并降低社会意愿偏差。而将列举实验分别作为自变量和因变量的示例分析表明, 相比直接提问方式, 列举实验对模型估计具有关键影响, 它可以提供较为可靠、可信和恰当的分析结果。当然, 列举实验仍然面临敏感问题测量的本质困难、实验效力、设计和分析难点等问题, 因此需要进一步优化改进。

关键词: 关键词: 列举实验; 调查实验; 敏感问题测量; 非制度化政治参与

中图分类号: C81 **文献标识码:** A **文章编号:** 1002 - 9753(2016) 09 - 0135 - 09

List Experiment in Measuring Sensitive Issue ——Example from Non-institutionalized Political Participation

WANG Pu-qu, JI Cheng-yuan

(School of Government, Peking University, Beijing 100871, China)

Abstract: It is difficult to measure sensitive issues in social science study, for respondents choose not to report or lie to the questions as a result of social desirable bias. With the development of survey technology, such a difficulty comes back to academic discussion in recent years, in which list experiment is three most popular methods. The key of list experiment design is to avoid design effect, ceiling effect and floor effect. This study takes non-institutionalized political participation as an example. We introduce the method to test the assumption. And we prove that list experiment can lower the nonresponse rate, raise the truthful response rate and lower the social desirable bias. Compared with direct question, list experiment has significant impacts on the model estimation, which provide more reliable, credible and appropriate result, when applying list experiment as dependent and independent. Of course, list experiment still has to face the problems include the essential problem of sensitive issues measuring, power, design and analysis, which means it still need further improvement.

Key words: list experiment; survey experiment; measuring sensitive issue; non-institutionalized political participation

收稿日期: 2016 - 03 - 19 修回日期: 2016 - 08 - 16

基金项目: 国家哲学社会科学重大项目“推进国家治理体系现代化研究”成果, 项目批准号(014ZDA011); 国家哲学社会科学基金后期资助项目“当代中国治理模式研究”成果。

作者简介: 王浦劬(1956—), 男, 江苏盐城人, 北京大学国家治理研究院院长, 北京大学政治发展与政府管理研究所所长, 教育部“长江学者”特聘教授, 北京大学政府管理学院教授。

“您是否有接受别人的礼物而帮助别人的经历?”“您是否同意外来务工人员应当享受与本地居民相同的福利待遇?”在社会调查中,对于这些直接的询问,受访者往往选择拒绝回答^[1],这将导致待测问题的低估,或者,受访者会隐藏自己的真实偏好或行为,而选择符合社会规范的答案,这将导致待测问题的社会意愿偏差(social desirable bias)。社会科学的调查研究中,敏感问题出现的概率很高,对于上述关于腐败的经历和社会歧视的问题,受访者会因为不合法或者存在社会意愿压力而拒绝应答或不作真实应答。在这种情形下,社会调查采集的调查数据不仅没有采集到受访者的真实想法,而且在进一步的统计推断和分析过程中可能产生错误的结论,从而严重影响实证研究的效度。

由此可见,敏感问题测量是实证研究的一个关键性方法论挑战,近年来,以普林斯顿大学政治学系为中心的国际社会科学界,掀起一股敏感问题方法论讨论的新热潮,在《政治分析》(*Political Analysis*)、《美国统计学杂志》(*Journal of the American Statistical Association*)、《美国政治学评论》(*American Political Science Review*)和《美国政治学杂志》(*American Journal of Political Science*)等期刊上,迅速推出一批研究成果。鉴于这一问题对于实证研究科学性的重要意义,本文首先概要介绍对于敏感问题的既往调查方法,分析其产生原因和相关调查技术的原理。之后,笔者以非制度化政治参与调查为典型案例,以示例分析和验证方式,检测了列举实验(list experiment)方法的有效性,并且分析了其需要进一步优化之处。

一、既往的调查方法

对于敏感问题的调查,常常产生受访者不应答或者虚假应答的情形。研究表明,之所以出现这种情况,其主要原因在于:社会调查的访问员无法细致观察和检验受访者在待测问题上的回答,而只能选择依靠受访者自报答案。在遇到敏感问题时,受访者的答案会直接暴露在访问员面前。在这种情况下,由于敏感问题多不合法或不符合社会规范,受访者将独自承担由此产生的道德、社

会甚至政治后果。由此产生的顾虑,往往会使得受访者选择回避应答或选择不真实应答。最初,研究者意识到敏感问题的社会意愿偏差后,仅仅是通过降低措辞的敏感性,期望由此获得更真实的应答。经验表明,这种方法效果乏善可陈。

经过几十年的发展,目前,在社会科学研究中的常用的应对敏感问题测量的方法一般包括随机回答技术(randomized response technique, RRT)、列举实验和支持实验(endorsement experiment)。在这其中,支持实验主要应用于(敏感)政治领导人的支持研究,例如伊斯兰武装^[2]和塔利班^[3]等。该方法将受访者对特定政策的支持和对特定领导人的支持组合起来,最终分离出对敏感领导人的支持。新近的研究成功地将分析策略颠倒,从支持实验中分离出对政策的真实支持程度^[4]。由于该方法的应用范围有限,因此不在本文的讨论范围之内。

随机回答技术的发明为调查研究中社会意愿偏差的解决提供了可行的办法,这其中最基本的方法是沃纳模型^[5]。按照沃纳模型,在受访者碰到需要随机回答的问题时,使用随机化装置(硬币、骰子等)进行一次随机实验,以决定是回答问题 A 还是回答问题 B,在这其中,问题 A 和问题 B 往往是相反的。例如问题 A 问“是否作弊过”,问题 B 就要求问“是否没有作弊过”。在这一过程中,受访者无需告知采访员随机化选择的具体结果和具体回答了哪个问题。假定回答问题 A 的概率为 p ,回答问题 B 的概率则为 $(1-p)$,因为两者是对立事件。其中 p 应当在访问前即已确定。那么,敏感选项的选择比例 $\hat{\pi}$ 可以用式(1)得出,其中 n 为样本总量, n_1 为回答“是”的比例。沃纳同时推导了标准差的计算公式。

$$\hat{\pi} = \frac{p-1}{2p-1} + \frac{n_1}{(2p-1)\pi} \quad (1)$$

沃纳的基本模型虽然效果比降低措辞的敏感性好,但是,该模型设置的 A 与 B 的相反问题,很容易引起受访人的注意,进而了解采访的真实意图,由此导致受访者完全可能采用针对性的回答方式。同时,式(1)表明 p 不能为 $1/2$ 。 $1/2$ 是一

个可以最大程度减少受访者顾虑的概率,无法设定为 1/2,也造成常见的随机化装置硬币等无效,这就需要采用其他更复杂的装置。

西蒙斯^[6]对沃纳模型进行了改进,构建了西蒙斯模型。这一模型将其中的问题 B 改成其他不相关(不敏感)的问题,降低了受访者的关注。同时,西蒙斯模型允许 p 直接等于 1/2。则敏感题 A 中选择“是”的比例 $\hat{\pi}_A$ 可由式(2)计算,其中 n 为样本量, n_1 为回答“是”的比例, π_B 为问题 B 回答“是”的比例(已知)。例如,如果是询问学生,问题 B 可以询问“您是高一年的学生吗?”,研究者可以事先确定调查对象中高一年学生的比例。西蒙斯模型在使用中需要了解非敏感的问题 B 的总体比例,访问者不能知道受访者对问题 B 的回答,否则将失去掩护作用。西蒙斯模型是应用较广的随机化回答技术。此后,不同学者提出更多改进策略,目前流行的随机化回答技术,是启动随机化装置(硬币等)后,受访者如果看到约定(例如硬币正面)的情况,直接机械式回答“是”或“否”中一种,因而概率为 0.5;约定的情况没出现,则真实回答敏感问题。相关研究参见 Blair *et al.*^[7]。这种方法取消了对问题 B 的限制,相当于直接将 π_B 定为 0.5,对问卷设计和执行都极为方便。

$$\hat{\pi}_A = \frac{\frac{n_1}{n} - (1-p)\pi_B}{p} \quad (2)$$

但是,以上这些随机化回答方法,仍然会让受访者注意到研究者和访问者的兴趣和意图所在,从而降低回答的真实度^[8]。这种对个体保护而对群体不保护的随机回答技术,仍然会让明知违法的受访者出于保护群体的目的而不作真实应答,因此,对于与整体利益相关的受访者,这种随机回答技术,在访问效果方面可能并无多大改善^[9]。这种技术在暴露研究关注点的同时,延长了访问时间。

二、列举实验及其设计

列举实验又称条目计数法(item count technique, ICT),由 Kuklinski 及其合作者提出^[10]。在 1991 年的美国种族和政治调查中,政治学者首

次设计了一道列举实验题^[11]以测量个人的种族歧视倾向。作为一种随机试验,研究者将受访者分为实验组和控制组两组。两组受访者分别需要回答如下问题:

【控制组】现在我将读出 3 件可能使你生气的东西。你只需告诉我其中____件使你生气,而不必说出具体哪项。

- (1) 联邦政府提高汽油税
- (2) 专业运动员薪水超过百万
- (3) 大型企业污染环境

【实验组】现在我将读出 4 件可能使你生气的东西。你只需告诉我其中____件使你生气,而不必说出具体哪项。

- (1) 联邦政府提高汽油税
- (2) 专业运动员薪水超过百万
- (3) 大型企业污染环境
- (4) 一个黑人家庭搬到你家隔壁

实验组和控制组的受访者都会面对 3 个控制选项,其中的区别在于,实验组的受访者将面对研究者关心的第 4 个敏感选项,即“一个黑人家庭搬到你家隔壁”是否使受访者生气。列举实验并不直接询问个人种族歧视倾向的敏感问题,而是转通过回答件数,降低了由于直接询问带来的社会意愿偏差。实际上,只要受访者不回答“4”,即所有事情都使受访者生气,研究者和访问员都无法确切了解该受访者的种族歧视倾向。这种信息的不对称,可以一定程度上鼓励受访者提供更真实的答案。

但是,列举实验题设计并不像看上去那么简单,它必须满足两个基本假定:

假定 1 不存在设计效应(design effect);

假定 2 不存在欺骗者,主要是指不存在天花板效应(ceiling effect)和地板效应(floor effect)。

所谓设计效应,是实验组增加的敏感选项,会对受访者选择控制题项产生显著的影响。设计效应会在受访者认定控制题项与敏感题项相关的情况下产生,会使得控制组和实验组的受访者面对控制题项时产生显著不同的回答方式。因此,列举实验题不存在设计效应,是列举实验题测量成

功的必要条件,而所谓列举实验题不存在设计效应,是指实验组增加的敏感选项,不会对受访者选择控制选项产生显著的影响。

列举实验与随机回答技术相比,理论上最大的缺点在于,随机回答技术仅看答案,没有任何人可以知道受访者的真实回答,而列举实验在一些情况下却无法达到这一点,这就会违背假定 2,即产生天花板效应或地板效应。所谓天花板效应,是指实验组的受访者面对的所有选项(J 个)都符合其实际情况,但是,为了保护隐私,受访者的应答变为($J-1$)个。所谓地板效应则正好相反,只有敏感的那个选项符合受访者的实际情况,但是,受访者宁可将答案变成 0,从而隐藏自己的偏好。

为了防止列举实验题失败,研究者在设计过程中,必须谨慎考虑设计效应、天花板效应和地板效应,控制题项的设置应当慎之又慎。一个关键技巧是控制题项之间应该至少有 2 个呈现强负相关关系,并且有一个应当会有大多数人选择^[12]。根据这一思路,询问个人的经历时,我们可以将出国经历、乘坐公共汽车的经历设置为控制选项。而干预题项的设置事关设计效应是否存在,应当选择指向清晰明确的干预题项^[13]。

三、列举实验的分析

列举实验题通过均值差来估计符合敏感题项状况的受访者比例,具体来说,假设控制组的平均值为 $\bar{X}_1$,对照组的平均值是 $\bar{X}_2$,那么,存在敏感题项状况的受访者总体比例为 $(\bar{X}_2 - \bar{X}_1)$ ^[14]。

但是,列举实验长期只能应用上述方法,通过比较控制组和实验组的均值差来估计敏感问题发生的人群比例,而无法进一步通过多变量回归分析探讨受访者的其他变量与敏感问题的关系。例如,具备哪些特征的受访者会隐瞒自己的种族歧视倾向。更进一步的问题是,如何将列举实验产生的变量作为我们熟悉的常规变量纳入多变量分析,Imai^[15]、Blair & Imai^[16]和 Imai *et al.*^[17]开发的多变量分析方法,开辟了这一路径。

下文将利用北京大学中国国情研究中心 2014 年设计并采集的“互联网与社交媒体调查”数据,对于这一方法进行典型实例分析和验证。该调查

采用该中心独创的“GPS 辅助的区域抽样”与分层多阶段相结合的抽样方法,对于解决传统户籍抽样中存在的难以覆盖流动人口、户籍资料不精确、人户分离等问题,具有重要积极作用。这项调查涉及 70 个县的初级抽样单位,共完成 3747 个个案,有效完成率为 66%。分析采用 R 程序的“list”包和“Amelia”包完成,前者应用于列举实验的分析,后者应用于多重插补。

非制度化政治参与,是政治社会学研究的热点敏感话题,研究者设计了一个列举实验题,想要了解参与“在请愿书上签名”和参与“游行/静坐/示威/群体性事件”的人群比例,问卷被分为 A、B、C 三种类型,列举实验题设计如下:

【A 卷】下面列举的事情,您做过几个? 您不必说具体哪个,请告诉我 0 到 3 之间的一个数字。

- a. 读大学
- b. 买房
- c. 换过工作

(出示答题卡) 您经历过上面提到的 ____ 件事情。

【B 卷】下面列举的事情,您做过几个? 您不必说具体哪个,请告诉我 0 到 4 之间的一个数字。

- a. 读大学
- b. 在请愿书上签名
- c. 买房
- d. 换过工作

(出示答题卡) 您经历过上面提到的 ____ 件事情。

【C 卷】下面列举的事情,您做过几个? 您不必说具体哪个,请告诉我 0 到 4 之间的一个数字。

- a. 读大学
- b. 游行/静坐/示威/群体性事件
- c. 买房
- d. 换过工作

(出示答题卡) 您经历过上面提到的 ____ 件事情。

A 卷为控制组, B 卷和 C 卷分别添加了敏感选项“在请愿书上签名”和“游行/静坐/示威/群体性事件”。同时,在列举实验后,又直接询问了受访者是否参与过“在请愿书上签名”和“游行/静坐/示威/群体性事件”,选项为“是”和“否”。当然,列举实验题和直接询问题中间,还设置有大量其他问题,尽可能减少其他干扰。

我们首先对列举实验题进行描述性统计和设计效应检验,结果见表1。直接提问的结果表明,无应答率为2.7%,而列举实验的无应答比率仅为0.7%,表明列举实验确实降低了无应答的比例。在列举实验中,从控制组、干预组1和干预组2的样本量看,大致均衡,分别有1223、1265、1259个样本,但是,无应答样本较少,不足1%,且控制组合干预组之间分布较平均。设计效应检验的原假定

是存在设计效应,我们需要比较邦费罗尼 p 值 (Bonferroni p) 与 0.05 的关系,如果前者大于 0.05,则拒绝原假设,认为不存在设计效应。我们的设计中,干预组1存在设计效应,干预组2则不存在显著的设计效应,因而对于“在请愿书上签名”列举实验测量失败。下文仅考察参与“游行/静坐/示威/群体性事件”的情况。

表1 列举实验的设计效应检验

取值	直接提问		控制组		干预组1 在请愿书上签名		干预组2 游行/静坐/示威/ 群体性事件	
	频数	百分比	频数	百分比	频数	百分比	频数	百分比
0	3591	95.8	308	25.2	377	29.8	326	25.9
1	54	1.4	583	47.7	519	41.0	524	41.6
2	—	—	259	21.2	279	22.1	319	25.3
3	—	—	67	5.5	72	5.7	76	6.0
4	—	—	—	—	8	0.6	3	0.2
无回答	102	2.7	6	0.5	10	0.8	11	0.9
样本量	3747		1223		1265		1259	
Bonferroni p			—		0.0099		0.6854	

排除了设计效应,我们还需要检验是否存在天花板效应和地板效应。首先,表2对控制题项有3个($J=3$),而存在天花板效应和地板效应的情形进行了举例。

表2 列举实验中存在天花板效应和地板效应的情况举例($J=3$)

取值	控制组	实验组
0	(0,0)	(0,0)(0,1)†
1	(1,0)	(1,0)(0,1)
2	(2,0)	(2,0)(1,1)
3	(3,0)	(3,0)(2,1)(3,1)†
4	—	(3,1)

注:(y,z)表示特定回答类型的人群。其中 y 为控制选项中回答“是”的个数, z 为对敏感选项的回答情况(是=1,否=0)。
†表示受天花板效应和地板效应影响的回答类型。

从表2看,(0,1)是可能受到地板效应影响的人群,只有敏感选项符合他们的实际情况,而为了

防止暴露真实想法,他们在回答列举实验题时,宁可应答“0”。(3,1)是可能受到天花板效应影响的人群,不仅3个控制选项符合情况,敏感选项也符合他们的情况,但是他们也选择应答“3”从而保护真实想法。

表3对我们的列举实验中每一种回答类型的人群比例进行了估算,并汇总了标准误。 $\hat{\pi}_{01}$ 和 $\hat{\pi}_{31}$ 就是可能存在地板效应和天花板效应的人群比例,分别为-0.81%和0.24%,这些数值非常小,表明我们的列举实验中并不存在严重的地板效应和天花板效应,模型分析中不需要进行相应的考虑。如果存在地板效应和天花板效应,模型设定中则需要对此进行校正,关于这一问题,可以参考Meng *et al* 的研究^[18]。

接下来,我们可以比较直接回答参与“游行/静坐/示威/群体性事件”的比例和列举实验表明参与“游行/静坐/示威/群体性事件”的比例,来确认非制度化政治参与经历的应答中受访者的社会

表 3 列举实验中不同回答类型的实际受访者比例估计

取值	$\hat{\pi}_{y0}(\%)$	标准误	$\hat{\pi}_{y1}(\%)$	标准误
0	26.12	0.012	-0.81	0.018
1	42.8	0.018	5.1	0.018
2	20.46	0.014	0.82	0.010
3	5.26	0.007	0.24	0.001
总计	94.64		5.35	

注： $\hat{\pi}_{yz}$ 为特定回答类型人群的估计比例。其中 y 为控制选项
中回答“是”的个数 z 为对敏感选项的真实回答（是 = 1，否 = 0）。

意愿偏差大小。分析中对于列举实验未回答的样本采用列删除方式排除，共 27 个，因为分析中无法容纳列举实验题为缺失值的情形。

表 4 非制度化政治参与经历的社会意愿偏差

非制度化 政治参与	直接提问	列举实验	社会意愿偏差
比例	0.0196	0.0783	0.0587
95% 置信区间	(0.0116 , 0.0276)	(0.0361 , 0.1204)	(0.0157 , 0.1016)

从表 4 的结果看，直接询问受访者是否参加过“游行/静坐/示威/群体性事件”，仅有 1.96% 的受访者确认有过这类经历。而使用列举实验，该比例提高到 7.83%，由此可见，真实应答的比例仅为四分之一。直接提问和列举实验的差值 5.87%，这个值即为受访者因为担心真实想法泄露等因素，或期望符合社会规范而产生的社会意愿偏差程度。这也表明，如果仅仅采用直接提问的方式，将严重低估激进政治参与的比例，由此产生的其他统计推断都会产生偏差。由此也可见，在问卷调查中，询问受访者激进政治参与的经历时，往往存在社会意愿偏差，由此需要应用列举实验方法来应对。

接下来，我们将采用示例方法，来验证列举实验题作为因变量和自变量的分析方法。针对列举实验题作为因变量的分析情形，Imai^[19] 和 Blair & Imai^[20] 基于前述 2 个假定，开发了基于非线性最小二乘法（nonlinear least squares，NLS）估计和极大似然估计的估计办法。极大似然估计的应用范围较广，但是估计的难度较大，因而估计失败的可能较大。我们的分析应用了非线性最小二乘法。

首先，列举实验题作为因变量时，我们举例考

察谁具有非制度化政治参与经历，或者说是具有非制度化政治参与经历的影响因素。变量描述性统计见表 5，其中男性受访者占 47%，平均 45.9 岁，平均接受 8.9 年教育，党员有 9.7%，有工作的受访者占 56.7%，体制内（党政机关和国有企事业单位）的受访者有 12.5%。收入水平为家庭收入状况自评 4 等级，非制度化政治参与倾向为回答“如果受到政府非常不公平待遇，或对政府的某个官员或某项政策非常不满，利用上网发帖的方式解决问题”选择“是”的比例，社会满意度利用对“社会稳定、经济形势、文化发展、公共秩序、法制改革、科技政策、环境保护、公民维权、食品安全”满意度制备，个人表达利用参与“向上级政府领导表达自己的观点、通过媒体表达自己的观点、通过社会组织表达自己的观点”的经历制备，政治价值观使用李程远等^[21] 文中的方法制备，得分从低到高对应个人的政治价值观由威权价值观到民主价值观，社会满意度、个人表达、政治价值观变量的制备均采用了项目反映理论。

由于存在较多的缺失值，以下分析中的结果均为 20 次插补数据分析基础上，利用鲁宾法则合并的结果，这样可以在比较中保证不同模型样本量一致，增加分析的可靠性。表 6 的结果呈现中，

表 5 变量描述性统计

变量	样本量	均值	标准差	最小值	最大值
性别(男 = 1)	3 720	0.470	0.499	0	1
年龄	3 720	45.854	16.173	18	89
受教育时间	3 639	8.939	4.561	0	24
收入水平	3 584	3.156	0.687	1	4
党员(是 = 1)	3 668	0.097	0.295	0	1
就业(是 = 1)	3 715	0.567	0.496	0	1
体制内(是 = 1)	3 480	0.125	0.331	0	1
非制度化政治参与倾向	3 697	0.132	0.339	0	1
社会满意度	3 656	0.000	0.930	-3.257	2.512
个人表达	3 668	-0.005	0.553	-0.237	2.265
政治价值观	3 462	0.564	0.116	0	1
非制度化政治参与经历	3 641	0.015	0.121	0	1
政治信任	3 501	0	1	-3.670	1.338

模型 1、模型 2、模型 3 的因变量为直接提问的非制度化政治参与,模型 4 和模型 5 的因变量则为列举实验所获的非制度化政治参与,模型 4、模型 5 的数据集使用了 A 卷和 C 卷数据,模型 3 也使用这一数据集,以防出现数据集差异减少带来的统计推断差异。模型 1 和模型 2 的结果表明,在控制了经济社会地位、体制羁绊因素和个人特质后,政治

价值观对于个人的非制度化政治参与经历居于负向影响,即更偏向威权价值观的个体,反而拥有更多的非制度化政治参与经历。但是,这一结果是真的吗?模型 4 和模型 5 的结果逆转了这一结果,在控制了其他因素后,个人的政治价值观偏向民主的受访者,更可能拥有非制度化政治参与经历。

表 6 模型结果摘要

变量	模型 1		模型 2		模型 3		模型 4		模型 5	
	系数	标准误	系数	标准误	系数	标准误	系数	标准误	系数	标准误
经济社会地位										
性别	0.997 ***	0.323	1.028 ***	0.324	1.177 ***	0.385	1.987 ***	0.787	2.298 ***	0.876
年龄	0.006	0.012	0.004	0.012	0.004	0.014	-0.035	0.026	-0.039	0.027
受教育时间	0.005	0.042	0.011	0.042	-0.092 *	0.051	-0.325 **	0.163	-0.370 ***	0.142
收入水平	-0.278	0.202	-0.304	0.202	0.177	0.250	0.563	0.722	0.567	0.549
体制羁绊										
党员	-0.249	0.448	-0.298	0.459	-0.158	0.522	-6.302 ***	2.292	-7.385 ***	2.329
就业	0.280	0.349	0.303	0.352	0.386	0.435	-1.938 **	0.943	-2.055 **	1.011
体制内	-0.494	0.469	-0.503	0.474	0.204	0.527	2.481 **	1.195	2.811 **	1.210
个人特质										
参与倾向	0.666 *	0.382	0.755 **	0.383	1.042 **	0.463	-2.951	5.270	-3.621	5.492
社会满意度	0.354 **	0.143	0.326 **	0.147	0.183	0.180	0.712 **	0.326	0.888 **	0.394
个人表达	1.094 ***	0.166	1.068 ***	0.168	1.087 ***	0.207	1.608 ***	0.553	1.840 ***	0.645
政治价值观			-2.540 **	1.244	-2.994 *	1.537			4.425 *	2.634
截距	-4.691 ***	0.995	-3.249 ***	1.201	-3.775 **	1.522	-1.093	2.190	-3.534	2.595
样本量	3720		3720		2465	2465		2465		

注: * p<0.1, ** p<0.05, *** p<0.01,下同。

这一结果表明,如果使用直接询问的方式,对敏感事项进行数据采集,并据此进行的统计推断,可能存在错误,列举实验的多元统计分析可以发现更加真实的结果。

列举实验题作为自变量的分析方法,直到最近才得以开发^[22],通过最大似然估计的两阶段法,可以方便地产生稳健估计,最大似然估计的应用,也拓展了方法的应用范围。我们以政治信任作为因变量,该变量为对“县市政府”和“中央政府”的信任程度因子得分。非制度化政治参与经历以及其他经济社会地位变量等为自变量,考察非制度化政治参与对于个人政治信任的影响。表 7 汇总了分析结果,模型 1 的非制度政治参与经历采用了

直接询问方式,而模型 2 采用的是列举实验获取的非制度政治参与经历。结果表明,在控制了其他因素后,直接提问和列举实验获得的非制度化政治参与经历,产生了明显差异,列举实验获得的非制度化政治参与经历对受访者的政治信任产生显著的负向影响。另外一个变化是,政治价值观对于政治信任的影响也消失了,这表明,个人的威权价值观或民主价值观,实际上并不会对政治信任产生显著影响。

综上验证可见,列举实验对于敏感问题测量具有重要价值,它可以有效化解一部分社会意愿偏差。更为关键的是,直接询问的测量方式可能对于统计推断产生显著的影响,这种影响甚至是

表 7 模型结果摘要

变量	模型 1		模型 2	
	系数	标准误	系数	标准误
经济社会地位				
性别	-0.121 **	0.037	-0.083 **	0.039
年龄	0.006 ***	0.001	0.007 ***	0.002
受教育时间	0.003	0.005	0.002	0.006
收入水平	0.124 ***	0.028	0.132 ***	0.028
体制羁绊				
党员	0.009	0.067	-0.021	0.077
就业	-0.064	0.043	-0.015	0.045
体制内	0.104	0.064	0.038	0.066
个人特质				
参与倾向	-0.065	0.059	-0.076	0.057
社会满意度	0.305 ***	0.022	0.277 ***	0.021
个人表达	0.234 ***	0.038	0.261 ***	0.040
政治价值观	-0.542 ***	0.169	-0.158	0.172
非制度政治参与经历	-0.230	0.148	-1.999 ***	0.158
截距	-0.320 **	0.163	-0.561 ***	0.168
样本量	2465		2465	

方向上的,我们关心的敏感变量实际上应该对因变量有正向影响,但是,直接提问的方式却可能产生负向的影响,由此可能蒙蔽研究者的眼睛。

四、结论与讨论

如何在调查中消解受访者的顾虑、如何更好地保护其真实想法、如何采集到更为真实的数据,实证研究的学者们一直为此而积极努力。目前,随机回答技术在社会科学各学科都已广泛应用,但是,主要由政治学者 Blair 和 Imai 推动的列举实验,尚属于新兴技术。本文的研究介绍了列举实验的设计、假定检验和分析路径,并利用相关全国概率调查数据进行了示例性验证。我们的验证表明,列举实验不仅能够提高应答率,而且能够有效地降低社会意愿偏差,获取较为真实的应答。并且在敏感问题的统计推断中也呈现关键价值,可以纠正模型的估计,使得实证分析结果较为可靠、可信和恰当。

当然,列举实验也并非已经完美。归结起来,其主要存在以下问题:

首先,列举实验仍然没有解决敏感问题测量

的本质,即提供受访者真实回答的动力。列举实验本质上是研究者以一定的信息牺牲,来换取受访者的信息补偿。但是,受访者仍然有足够多的顾虑避免自己惹麻烦。文化程度较高的受访者可以轻松猜中研究者和访问者的意图,文化程度较低的受访者则可能对于这种询问方式非常困惑。事实上,在敏感问题的社会科学研究中,投入大量时间建立的个人信任,被证明往往对真实的分析与结果至关重要,例如著名的《街角社会》就是如此。因此,在一定程度上可以认为,对于敏感问题的大规模测量,一些分析工具和方法形成的结果与事实之间仍然有一段距离,为此,研究者对于这一方法的采用必须保持充分的谨慎和敬畏。

其次,列举实验问题的效力有待进一步检测。根据 Mulders *et al.* 的总结^[23],即使是 1965 年沃纳开发的随机回答技术,也只有 5 篇相关论文比较研究了这一技术在促进测量接近真实的程度问题。而研究者使用列举实验方法,究竟能在多大程度上接近真实,即列举实验的结果与事实之间究竟有多大差距,目前尚缺少大规模的检验。实际上,只有学者 Rosenfeld *et al.* 进行了初步讨论^[24],因此,列举实验问题的效力有待进一步检测。

再次,列举实验的问题设计需要精确化。这是列举实验测量成功与否的关键,研究者需要仔细考虑控制题项的设置,以保证分析假定的成立,保证不存在设计效应、天花板效应和地板效应。为了检测设计的可靠性,在正式调查前,对于列举实验设计进行小规模的测试至关重要。

最后,列举实验的分析路径和技术有待优化。由于实验组和控制组的设计、估计方法复杂性,列举实验需要更大的样本量,估计方法的复杂性也造成了分析结果的不稳定。更为关键的是,目前,列举实验与因果推断(Casual Inference)的结合还缺乏特别可行的分析路径,有待进一步的研究。

参考文献:

[1]严洁,任莉颖. 政治敏感问题无回答的处理:多重插补法的应用[J]. 华中师范大学学报:人文社会科学版, 2010(2): 29-34.
[2] Bullock W, Imai K, Shapiro J N. Statistical analysis of endorsement experiments: Measuring support for militant

- groups in Pakistan [J]. Political Analysis, 2011, 19(4): 363-384.
- [3] Lyall J, Blair G, Imai K. Explaining support for combatants during wartime: A survey experiment in Afghanistan [J]. American Political Science Review, 2013, 107(04): 679-705.
- [4] Rosenfeld B, Imai K, Shapiro J N. An empirical validation study of popular survey methodologies for sensitive questions [J]. American Journal of Political Science, 2015.
- [5] Warner S L. Randomized response: A survey technique for eliminating evasive answer bias [J]. Journal of the American Statistical Association, 1965, 60(309): 63-69.
- [6] Greenberg B G, Abul-Ela A L A, Simmons W R, et al. The unrelated question randomized response model: Theoretical framework [J]. Journal of the American Statistical Association, 1969, 64(326): 520-539.
- [7] Blair G, Imai K, Zhou Y Y. Design and analysis of the randomized response technique [J]. Journal of the American Statistical Association, 2015, 110(511): 1304-1319.
- [8] Droitcour J, Caspar R A, Hubbard M L, et al. The item count technique as a method of indirect questioning: A review of its development and a case study application [J]. Measurement errors in surveys, 1991, 11: 185-210.
- [9] 纪竹荪. 随机化回答技术调查“敏感性问题”的博弈分析及设计改进——对一次沃纳方法调查实践的分析思考 [J]. 统计与信息论坛, 2008(5): 82-86.
- [9] Kuklinski J H, Cobb M D, Gilens M. Racial attitudes and the “New South” [J]. The Journal of Politics, 1997, 59(02): 323-349.
- [11] Sniderman, P. M., P. E. Tetlock, and T. Piazza. Codebook for the 1991 National race and politics survey [R]. Survey Research Center, Berkeley, California.
- [12] Corstange D. Sensitive questions, truthful answers? Modeling the list experiment with listit [J]. Political Analysis, 2009, 17(1): 45-63.
- [13] Blair G, Imai K. Statistical analysis of list experiments [J]. Political Analysis, 2012, 20(1): 47-77.
- [14] Meng T, Pan J, Yang P. Conditional receptivity to citizen participation evidence from a survey experiment in China [J]. Comparative Political Studies, 2014: 0010414014556212.
- [15] Imai K. Multivariate regression analysis for the item count technique [J]. Journal of the American Statistical Association, 2011, 106(494): 407-416.
- [16] Blair G, Imai K. Statistical analysis of list experiments [J]. Political Analysis, 2012, 20(1): 47-77.
- [17] Blair G, Imai K, Zhou Y Y. Design and analysis of the randomized response technique [J]. Journal of the American Statistical Association, 2015, 110(511): 1304-1319.
- [18] Meng T, Pan J, Yang P. Conditional receptivity to citizen participation evidence from a survey experiment in China [J]. Comparative Political Studies, 2014: 0010414014556212.
- [19] Imai K. Multivariate regression analysis for the item count technique [J]. Journal of the American Statistical Association, 2011, 106(494): 407-416.
- [20] Blair G, Imai K. Statistical analysis of list experiments [J]. Political Analysis, 2012, 20(1): 47-77.
- [21] 季程远, 王衡, 顾昕. 政治价值观与网络抗争行为的限度 [J]. 社会, 2016.
- [22] Blair G, Imai K, Zhou Y Y. Design and analysis of the randomized response technique [J]. Journal of the American Statistical Association, 2015, 110(511): 1304-1319.
- [23] Lensvelt-Mulders G J L M, Hox J J, Van Der Heijden P G M. How to improve the efficiency of randomised response designs [J]. Quality and Quantity, 2005, 39(3): 253-265.
- [24] Rosenfeld B, Imai K, Shapiro J N. An empirical validation study of popular survey methodologies for sensitive questions [J]. American Journal of Political Science, 2015.

(本文责编: 海洋)