

公共政策的效果能否被准确预测?

来自中国塑料袋使用限制政策的自然实验证据

何浩然*

摘 要 本研究以塑料袋使用限制政策的实行为契机,运用自然实验方法对采用陈述性偏好法(stated preference methods)预测政策效果的外部有效性(external validity)进行检验。在该政策执行前塑料袋均免费时,运用陈述性偏好法调查预测该政策的效果,即在塑料袋被收费的假设情景下消费者对其的预测消费量;在该政策执行后塑料袋实际收费时,再次调查该政策的实际效果,即消费者的实际塑料袋消费量。因此获得了独特的数据以对预测和实际的政策效果进行比较分析。研究表明,运用陈述性偏好法能对实际政策效果提供准确预测。

关键词 陈述性偏好,外部有效性,公共政策

一、引 言

诸如税收、价格、移民,及社会保障等许多公共政策的设计和选择都常依赖于之前对政策效果的预测。由于公共政策的实行多具有难以逆转性,且不适政策的实行会带来较高负面成本,故在政策实行前准确预测政策可能产生的效果十分重要。能够越准确地预测特定政策设计的效果,则就越可能设计出或从一系列可选方案中选出合适的政策。因此,如何准确预测政策成为公共政策研究的热点领域。

如果相似的政策曾在别处被实行,可借鉴其之前经验来预测在本处实行该政策的效果。然而如果将要实行的是新政策或别处之前的政策实行经验并不完整,则主要可用两种方法对政策效果进行预测——调查方法和实验方法。

* 北京师范大学经济与工商管理学院、瑞典哥德堡大学经济系。通信地址:北京市海淀区新街口外大街19号北京师范大学后主楼16层,100875;电话:(010)58807847;E-mail:haoran.he@bnu.edu.cn, haoran.he@economics.gu.se。作者感谢翁茜、Fredrik Carlsson、Michael Hanemann、Mitesh Kataria、Peter Martinsson,以及在哥德堡大学、北京大学和中国人民学院的学术研讨会的参加者在本研究的设计和成文过程中所给予的有益建议,感谢贵州大学和北京大学的调查队在数据收集阶段所提供的帮助。感谢北京师范大学青年教师社会科学基金项目(105507GK)、环保公益性行业科研专项(201009066)、国家自然科学基金青年项目(11CJY037)、北京师范大学985工程(2011年度)三期平台建设项目,以及瑞典海外发展研究署发展中国家能力建设项目建设对本研究的资助。作者文责自负。

陈述性偏好 (stated preference) 调查方法是被广泛应用于对管理非市场物品或公共物品的政策的实行或变化效果进行预测的方法,其特点是能够准确描述可能发生的政策变化后情景并直接从受影响群体处收集其在该情景下的陈述性反应。然而,该领域文献对于受访者在调查中假设情景下所预测的陈述性反应(所说)与其在实际情况下的实际行为反应(所做)之间是否存在差异(也即该方法预测结果的外部有效性)的讨论仍未达成一致。另一方面,实验方法虽然能为受访者提供金钱激励并直接观察其在实验中表现出的行为反应,却也存在其缺点。例如,实验室实验由于难以在实验室构建出与实际政策环境较相近的实验情景,其结果可能存在更严重的外部有效性问题(Levitt and List, 2007);而实地实验(field experiment)由于构建难度大和成本高等局限,可以构建来预测政策效果的机会相当有限(Burtless, 1995)。本文将我国塑料袋使用限制政策(以下简称“限塑政策”)的执行作为自然实验,在其创造出的执行前和执行后这两个自然实验局中对同一采样池各进行一次调查:在政策前运用陈述性偏好法获得消费者在政策执行后塑料袋收费的假设情景下塑料袋消费的条件价值行为(contingent behavior),也即对政策后的消费行为的预测;在政策后则直接观测并记录消费者对收费塑料袋的实际消费行为。通过对两种行为的比较分析来检验陈述性偏好法对政策效果预测的外部有效性,并对影响该外部有效性的相关因素进行探讨。

中国政府自2008年6月起实行一项以减少塑料购物袋¹使用为目的的全国性公共环境政策。这一政策的核心措施是强制所有商品零售商对其之前免费提供塑料袋实行自主定价收费,且规定所定的塑料袋价格不得低于其经营成本(中华人民共和国商务部等,2008;中华人民共和国国家发展和改革委员会,2008)。尽管已有一些国家实行了与中国相似的基于市场的限塑政策,但其效果在各国差异显著:丹麦和爱尔兰等发达国家的限塑政策有效遏制了塑料袋的使用(Danish Environmental Protection Agency, 1999; Convery 等, 2007),但类似政策在南非等发展中国家却很快失去了效果(Hasson 等, 2007; Dikgang 等, 2012)。因此中国限塑政策的效果难以根据他国经验进行预测。鉴于在本研究的前后两次调查期间除该政策被执行外,与塑料袋消费相关的外部经济环境并未出现明显变化且未开展与限塑减塑相关的宣传活动,故消费者塑料袋消费量的变化应完全归因于该限塑政策的执行。²

相关文献对陈述性偏好法预测的外部有效性曾进行过一定考察,但极少有研究能以实际发生的政策变化为契机来对这一问题进行深入探讨。文献中

¹ “塑料购物袋”在下文中多被简称为“塑料袋”。

² 从部分所调查超市获得的2007—2008年的24个月的月销售收入和两类塑料袋的月消耗数据的变化趋势也可对此予以证明:该政策前免费政策后收费的塑料购物袋的消耗量在政策执行前和执行后也均分别保持稳定,该类塑料袋的月消耗量仅在政策执行当月(2008年6月)出现大幅减少,而月销售收入和政策前后均免费的塑料预包装袋(连卷袋)的消耗量在24个月内保持稳定。详细信息见He(2012)。

对该方法预测的外部有效性的检验主要有两种方法：一种方法是以显示性偏好 (revealed preference) 法³获得的估计为基准来检验陈述性偏好法预测的外部有效性 (Carson 等, 1996)。但由于显示性偏好法的估计同样存在误差 (见, 如 Mitchell and Carson, 1989), 以这样的估计为基准来进行检验, 其检验结果的有效性值得商榷。另一种方法则是以在实验室实验或实地试验中观测到的行为为基准来检验陈述性偏好法预测的外部有效性 (如 Shogren 等, 1999; Lusk 等, 2006; Chang 等, 2009)。但此类实验研究都是通过将消费者对各种全新商品 (如辐照鸡肉, 不含抗生素猪肉, 以及有机食品等) 的购买行为预测与其在实验中的购买行为比较来检验该预测的外部有效性。并且研究者在实验中都试图对引入的新商品进行某种方式的强调——如放置于显著位置, 进行介绍或提示等——来吸引消费者注意, 从而可能致使消费者在实验中较多购买所用来进行检验的新商品。而且这些研究多是用新商品在单个杂货店和较短时间段内的购买行为作为检验基准, 可能导致检验结果的普适性较差。因此, 从更长的时间段和更多类型的商店获得数据来进行比较应能提高检验结果的可信度。然而, 由于难以找到实际发生的政策变化机会来检验预测的外部有效性, 相关文献中仅有 Lampi and Orth (2009) 曾利用一项公共政策的变化来检验用陈述性偏好法对博物馆访问者类型变化预测结果的外部有效性。

可能降低陈述性偏好法对政策效果预测的外部有效性的因素存在客观和主观两方面。客观方面是由于受访者对政策前调查中给出的假设情景不熟悉, 因而难以想象或描述自己在该情景下的偏好, 导致其所进行的行为预测含有猜测的成分。受访者在政策后的实际偏好和行为因其对实际情况的逐渐熟悉而与之前的含有猜测的预测产生区别。主观方面则是由于受访者存在如影响政策内容或政策实行方式等动机而故意采取策略性回答, 即受访者在政策前调查中故意隐瞒自己真实的偏好和行为。为降低客观因素的影响, 本研究采用塑料袋这一大家都极为熟悉的日常消费品作为研究媒介, 鉴于绝大多数受访者都应十分清楚塑料袋对自己的价值, 从而剔除受访者由于不熟悉而进行的猜测。为降低主观因素的影响, 本研究的政策前调查被安排到距离政策执行日很接近时进行, 调查前该政策的内容、执行方式和执行日期等信息都已被确定并被公之于众, 从而剔除受访者的策略性回答动机。另外, 由于本研究在不同城市和不同类型商店对消费者进行大规模调查, 样本涵盖具有广泛社会经济特征的各类人群, 因而保证了本研究对陈述性偏好法外部有效性的检验结果具有较高可信度。最后, 我们使用塑料袋这一众人皆知的环境有害品为媒介来进行外部有效性检验, 还为了能同时考虑到在许多运用陈述性偏好法估计非市场物品价值的研究中均存在的受访者可能希望表现得较为环境

³ 显示性偏好法包括怡值定价 (hedonic pricing) 法、旅行成本 (travel cost) 法、趋避行为 (averting behavior) 法、疾病成本 (cost of illness) 法等。

友好的这一动机。

本文研究结果表明,所用陈述性偏好估计技术能够准确预测该公共政策的实际效果,从而也增强了运用陈述性偏好调查方法对政策效果进行预测的信心。本文余下部分按以下结构展开:第二部分介绍实验设计和检验方法;第三部分对所得数据进行分析;第四部分报告检验结果;第五部分总结和讨论本研究的发现。

二、研究方法

(一) 实验设计

本研究采用拦截式调查方式在政策执行前后各进行一次调查:调查员在商店出口处随机对已完成购物的消费者进行访问,记录该消费者本人在调查当次的单次购物中塑料袋消费量及其个人和家庭的相关信息。本研究采用受试者间(between-subject)实验设计,并保证前后两次调查在同样商店和同样时间段进行抽样从而获得可比较的研究样本。采用受试者间设计能够避免在前后两次调查中均采访同一位受访者而可能产生的记忆效应,并且也能够较容易地在更大范围进行前后两次调查以获得更具代表性的样本。

政策前的调查等到据政策执行约一个月前进行,其时该政策的内容和执行日期等信息已被多数民众所熟知,⁴从而使受访者能较容易理解和回答调查中的问题,同时也剔除了策略性回答动机。政策前调查除了记录受访者当次购物的免费塑料袋消费量并收集受访者及其家庭的相关社会经济信息外,主要是通过询问一系列陈述性偏好问题以获得塑料袋在被收取不同特定价格的假设情景下该受访者塑料袋消费的条件价值行为,即在各价格下所预测的塑料袋消费量。由于在预调查中发现受访者更容易从塑料袋收费后自己会比免费时减少多少比例的消费量的角度来进行考虑,故在政策前调查的陈述性偏好问题中,我们询问受访者在各特定假设价格下其预测会比收费前减少的消费量百分比,然后再通过“预测消费量=实际消费量×(1-预测消费量减小百分比)”的公式来计算其预测的塑料袋消费量。该政策允许商品零售商以不低于塑料袋经营成本为原则自主设定价格。我们在政策前咨询了塑料袋生产行业、零售业和政府相关部门的专家后,⁵所得信息表明最可能的塑料袋平均

⁴ 80%以上的政策前受访者报告他们在参加该调查前即已知道这项政策。值得注意的是,我们在对消费者和商店经理进行访谈后没有发现任何证据显示消费者的塑料袋消费行为由于发布政策即将实行的消息而已发生变化,而从部分调查超市获得的月度塑料袋消耗数据也支持同样的结论。详细讨论见 He (2012)。

⁵ 我们考虑了包括塑料袋的生产成本和获得成本,政府各相关部门对实行该政策的准备,商品零售商对执行该政策以及对从收费中牟利的态度等所有政策前可能获得的相关信息。

价格应在 0.5 元左右。考虑到不同商店自主设定的塑料袋价格可能存在区别,除 0.5 元外,我们也将 0.3 元和 1 元作为可能价格在政策前调查的陈述性偏好部分进行询问,从而能够获得塑料袋消费量在该价格区间内的预测变化。调查员在陈述性偏好部分首先会让受访者预测其在 0.5 元价格下塑料袋消费的条件价值行为,之后根据所答再让其预测在 0.3 元或/和 0.5 元价格下相应的条件价值行为。图 1 显示陈述性偏好问题的结构。如图 1 所示,受访者可能产生 11 种类型的反应,而其中 2 种反应是偏好不一致的,故在图中用十字叉标示。

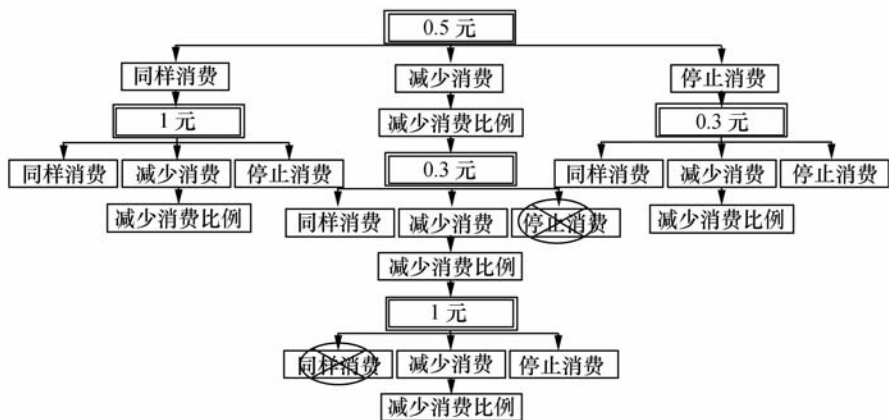


图 1 政策前问卷中陈述性偏好问题的结构

政策后的调查则在政策执行后四个月左右进行,那时消费者已获得了足够的时间去适应该政策的影响,从而所表现出的塑料袋消费行为已趋于稳定。政策后调查收集了受访者其时的塑料袋消费实际行为及其为所消费的塑料袋所支付的平均价格,以及与政策前调查相同的社会经济信息。政策后调查与政策前调查在同样的商店和同样的时间段开展。在政策执行后,政策前调查所估计的三个价格中有两个成为了常见的塑料袋实际平均价格,但却极少有塑料袋被收取 1 元/袋的价格。⁶因此本文检验陈述性偏好法预测的外部有效性仅基于 0.3 元和 0.5 元价格下的相关信息进行。

本研究的目标是对消费者政策前预测的消费行为和政策后发生的实际行为进行比较并检验两者是否存在差异,故所需检验的研究假设为:在相同价格下,消费者政策前预测的塑料袋消费量(Q_p^P)等于政策后实际消费量(Q_p^A)。以 0.3 元价格下的情况为例,原假设即为: $Q_{p=0.3}^{A\text{total}} = Q_{p=0.3}^{P\text{total}}$ 。本文将在 0.3 元和 0.5 元价格下对该原假设分别进行相应的统计检验。此外,由于政策

⁶ 我们意识到可能存在由于消费者自选择去不同类型商店购物从而面临不同塑料袋价格的可能性,但同时也注意到由于其价格在各商店的差别很小而不足以影响消费者选择所访问商店的决策。因此我们对面临不同塑料袋平均价格的受访者的社会经济特征进行检验。相关统计检验的结果表明我们不能在 5% 的统计显著水平上拒绝大多数社会经济特征在面临塑料袋实际价格为 0.3 元和 0.5 元的两组消费者之间不存在差异的原假设(见表 5)。

的不完全执行可能会对预测的外部有效性产生影响,我们还将进一步构建计量模型,在控制政策执行情况影响的基础上,对该原假设进行多元(multivariate)检验。上述研究方法系沿用 He (2010) 所发展的方法。

为获得具有良好代表性的研究样本,政策前后的两次调查都在不同城市,不同居民区,不同类型的商品零售场所,工作日和节假日,以及一天中的不同时间段开展。两次调查均在北京和贵阳两个城市开展,覆盖了发达地区和欠发达地区的两类典型城市。调查商店类型选择了超市和农贸市场这两类普通消费者访问最为频繁的商品零售场所。另外调查覆盖了工作日和节假日,以及在每个调查日覆盖了所有购物时段。调查最终共获得 3 077 个⁷有效样本。如表 1 所示,我们尽可能让样本的空间和时间来源分布均匀,从而保证样本的代表性以及政策前后样本的可比性。

表 1 两次调查的样本的空间和时间分布

调查时段	北京样本		贵阳样本		所有地区 和市场
	超市	农贸市场	超市	农贸市场	
7:30—11:00	226	202	276	285	989
12:00—15:00	196	194	348	272	1 010
17:30—20:00	202	190	281	405	1 078
所有时段	624	586	905	962	3 077

注:三个调查时段覆盖了所调查商店的所有主要购物时段。

(二) 多元检验方法

鉴于本研究所采用的是混合截面数据,为了在控制其他影响因素的基础上检验陈述性偏好法在政策前预测的塑料袋消费行为与消费者在政策后的实际行为的一致程度,我们构建可将政策前后面临同样假设价格和实际价格的受访者样本分别包含在内的普通最小二乘(OLS)计量模型来进行多元检验。⁸具体而言,该模型的因变量包含了政策前预测的在特定假设价格下单次购物塑料袋消费量(Q_p^p)和政策后实际观测到的在同一实际价格下单次购物塑料袋消费量。由于通过计算获得的 Q_p^p 是政策前免费塑料袋实际消费量($Q_{p=0}^A$)的分数,因而是一个连续变量。模型的自变量向量 $\mathbf{X}=(\mathbf{X}_0, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_m, \mathbf{X}_n, \mathbf{X}_r)$ 包含一系列变量。其中, \mathbf{X}_i 为标示本研究关键变量的受访者是否面临实际价格的二元虚拟变量“实际价格”,而引入其他自变量的目的则主要是为了在多元检验中控制其他因素的影响: \mathbf{X}_j 表示受访者所消费的全部塑料袋中

⁷ 15 个受访者样本由于单次购物塑料袋消费量极高而被作为不能代表普通消费者塑料袋消费行为的异常值被剔除。

⁸ 由于因变量数据具有分位数(quantile)特征,我们也尝试过运用分位数回归模型对相同变量设定的模型进行估计,发现回归结果在因变量各分位区间的区别很小,表明不存在使用分位数回归模型的必要性。为解释方便起见,本文的讨论仍基于普通最小二乘模型的回归结果。

收费塑料袋所占的比例⁹，以此来控制政策执行力度的影响； X_m 包含体现受访者对该政策的认识以及对不使用塑料袋所持态度等方面的变量； X_n 包含受访者本人及其家庭的社会经济变量； X_r 则包含一系列控制所开展调查的地区、商店类型、工作日或节假日，以及不同时间段等对塑料袋消费行为可能造成影响的变量； X_0 表示该模型的截距项。此外，我们还在计量模型中引入了相关变量的二次项和交叉变量。本文的数据描述部分将对各变量进行详细说明。

在多元检验控制住其他可能对塑料袋消费行为具有影响的因素后，“实际价格”虚拟变量的系数表明消费者的塑料袋消费行为受到实际价格与受到同样的假设价格的影响是否存在差异。如果实际价格的影响等同于同样的假设价格的影响，也即消费者在政策后面临实际价格时所消费的塑料袋与其在政策前陈述性偏好调查中面临同样假设价格时所预测消费的塑料袋一样多，则这一“实际价格”变量的系数应在统计上不显著异于零。否则，这一系数显著大于（或小于）零则表示消费者在政策前调查中低估（或高估）了其在政策后相同价格下的实际塑料袋消费量。同样的多元检验模型将被用于对0.3元和0.5元两种价格下的预测进行检验。

三、数据描述

（一）政策前的预测消费量与政策后的实际消费量比较

政策前调查的1041位受访者中仅有13位在调查当次的单次购物中未消费免费塑料袋。图1所示的11种预测反应类型中有两种类型的反应为偏好不一致，政策前调查的所有1041位受访者的预测反应中均未出现属于这两种不一致偏好的类型。此外，1041位受访者中有13位给出了价格越高（或低）将消费越多（或少）的塑料袋的偏好不一致的预测。如此小的偏好不一致比例同时表明受访者认真回答了该陈述性偏好调查问卷。在本文后面的分析中，我们将去掉这13个偏好不一致的样本而基于余下的1028个样本的数据进行分析。表2数据显示，具有最常见的三种预测反应类型的受访者比例超过82%，而所有这三种反应类型均体现出塑料袋消费量会随收费政策的完全执行而大幅下降。

表2 对假设价格产生不同类型预测反应的受访者所占比例

预测反应类型	0.5元	0.3元	1元	频数(观察值数)	百分比(%)
1	同样消费	n. a.	同样消费	32	3.11%
2	同样消费	n. a.	减少消费	30	2.92%
3	同样消费	n. a.	停止消费	9	0.88%

⁹ 为了降低受访者在政策前调查中进行预测的难度，在陈述偏好问题中的假设情景被描述为“所有商店都收费”，也即意味着上述比例在预测消费量的所有样本中均为100%。

(续表)

预测反应类型	0.5元	0.3元	1元	频数(观察值数)	百分比(%)
4	减少消费	同样消费	n. a.	37	3.60%
5	减少消费	减少消费	同样消费	0	0.00%
6	减少消费	减少消费	减少消费	293	28.50%
7	减少消费	减少消费	停止消费	270	26.26%
8	减少消费	停止消费	n. a.	0	0.00%
9	停止消费	同样消费	n. a.	4	0.39%
10	停止消费	减少消费	n. a.	72	7.00%
11	停止消费	停止消费	n. a.	281	27.33%
直接被询问该假设价格的观察值数	1 028	957	634	1 028	100%

注:(1)第5和第8项预测反应类型为不一致偏好;(2)“n. a.”表示该受访者在政策前陈述性偏好调查中未被要求在该假设价格下预测其反应行为。

由于受访者在政策前调查的陈述性偏好部分中会先对其在0.5元假设价格下的反应行为进行预测,然后根据其所预测的行为再被要求对0.3元或/和1元假设价格下的行为进行预测(如图1所示)。故所有1028位受访者均对其在0.5元假设价格下的行为进行了预测,却分别有71和394位受访者未对其在0.3元和1元假设价格下的行为进行预测。在满足塑料袋消费量随价格升高而递减¹⁰以及个人偏好保持一致(*consistent preferences*)两个前提条件下,我们能够构建以下两条推断规则,分别对全部71位未对0.3元和394位中的357位未对1元假设价格下的反应进行预测的受访者的预测反应信息进行补充:第一,若某位受访者预测在特定假设价格下将会消费同免费时同样多的塑料袋,则推断其在较低假设价格下也将会消费同免费时同样多的塑料袋;第二,若某位受访者预测在特定假设价格下将会完全停止消费塑料袋,则推断其在较高假设价格下也将会完全停止消费塑料袋。本文后面的分析将基于补充了完整信息的数据进行。

表3报告了以样本观察值频数和累计百分比衡量的在各假设价格下受访者塑料袋消费的条件价值行为。其数据表明,受访者预测其塑料袋消费量将随价格上升而不断减少:当价格从零上升至0.3元至0.5元再至1元时,越来越多的受访者选择不断减少塑料袋消费直至最终停止塑料袋消费。在0.3元、0.5元和1元假设价格下,预测自己将会停止消费塑料袋的受访者比例分别为27.4%、34.7%和64.1%,而预测自己将会消费同免费时一样多塑料袋的受访者比例则分别为10.7%、6.8%和3.1%。通过表3数据计算出的加权平均结果还表明,在政策被完全执行且塑料袋价格为0.3元、0.5元和1元的假设情景下,受访者预测将减少的塑料袋消费量百分比分别为63.2%、70.5%和88.3%。

¹⁰ 表2所示的三种最常见的预测反应均表明塑料袋消费量与其价格呈显著负相关关系。

表3 补充信息^a后对假设价格产生不同类型预测反应的受访者的频数和所占比例

Responses	0.3元		0.5元		1元	
	观察值	累计	观察值	累计	观察值	累计
	频数	百分比	频数	百分比	频数	百分比
The same use	112	10.7%	71	6.8%	32	3.1%
Reduce use 0%—20%	19	12.6%	11	8.0%	7	3.9%
Reduce use 20%—40%	102	22.5%	64	14.1%	11	5.1%
Reduce use 40%—60%	235	45.4%	217	35.2%	53	10.3%
Reduce use 60%—80%	168	61.8%	191	53.9%	91	19.6%
Reduce use 80%—100%	111	72.6%	117	65.3%	161	35.9%
Stop use	281	100.0%	357	100.0%	636	100.0%
各价格下的观察值数	1 028		1 028		991	

注：“补充信息”指运用上文所说明的信息补充规则将受访者在某些价格下未直接被要求预测的反应行为信息补充完整。

政策后调查的2036位受访者¹¹中有889位受访者在调查当次的单次购物中未使用新塑料袋，表明该政策促使相当比例的消费者完全停止消费塑料袋。鉴于中国零售行业激烈竞争的市场环境，各类商店均尽可能地调低了塑料袋价格：塑料袋的实际价格在0.2—0.6元的区间内波动，因城市、居民区和商店类型不同而有所差别。在本研究调查所涉及的超市和农贸市场，99%的受访者报告塑料袋平均实际价格接近0.3元或0.5元；而从两个开展调查城市的所有商店看，最常见的平均价格也约为0.3元或0.5元，但极少接近1元。因此本研究对陈述性偏好法预测的外部有效性检验仅基于0.3元和0.5元价格下的数据进行。

表4数据显示受访者的单次购物塑料袋消费量在各价格下的政策前预测均值和政策后实际均值。当政策前实际价格为零时，受访者的单次购物塑料袋实际消费量为2.97袋/次；而当价格在陈述性偏好问卷的假设情景下升至0.3元至0.5元再至1元时，受访者预测的单次购物塑料袋消费量均值分别下降至1.06、0.85和0.32袋/次。当政策执行后塑料袋实际收费时，调查所记录的数据显示在0.3元和0.5元的平均价格下，单次购物塑料袋实际消费量均值分别为1.03和0.95袋/次。在相同价格下，预测消费量与实际消费量相当接近。表4最后一列数据报告了在各价格下预测和实际完全停止消费塑料袋的受访者所占比例，可以看到在相同价格下，实际完全停止消费塑料袋的受访者比例均大于对应的预测值。

¹¹ 由于在政策前调查中每位受访者均能在不同价格的假设情景下对自己的反应行为进行预测，而政策后每位受访者的实际反应行为却只可能基于其所面临的那一个实际平均价格而产生，故在政策后调查中抽取了更大规模的样本。此外，2036位政策后受访者中的104位由于不能指出其所面临的塑料袋实际平均价格，故未被纳入本研究检验结果部分的分析数据。

表4 单次购物塑料袋消费量的统计描述

	单次购物中 塑料袋消费	观察值数	均值	标准差	停止消费塑料袋 受访者所占比例
政策前调查所获得的 实际和预测的塑料袋 消费量	$Q_{p=0}^A$	1028	2.97	1.829	1.26%
	$Q_{p=0.3}^P$	1028	1.06	1.237	28.21%
	$Q_{p=0.5}^P$	1028	0.85	1.085	35.60%
	$Q_{p=1}^P$	991	0.32	0.669	64.58%
政策后调查所获得的 实际的塑料袋消费量	$Q_{p=0.3}^A$	1594	1.03	1.246	41.34%
	$Q_{p=0.5}^A$	338	0.95	1.313	48.52%

注： Q_p^P 表示在假设价格 p 下预测的塑料袋消费量； Q_p^A 表示在实际价格 p 下实际发生的塑料袋消费量。

(二) 影响因素统计描述

对各潜在影响因素的统计描述按不同的子样本分组列于表5。列[1]显示政策前调查的1028位面临0.3元和0.5元假设价格的受访者样本，而列[2]和列[3]则分别反映政策后调查的1594和338位面临0.3元和0.5元实际平均价格的受访者样本。列[4]则表示两次调查的所有样本。所有大中型超市基本都完全执行了对塑料袋收费的政策，而以农贸市场为代表的其他类型商店却没有完全执行。政策执行后消费者所使用的全部塑料袋中收费塑料袋所占的比例仅为40%，表明政策执行远没有达到令人满意的程度。所有其他变量在各子样本组别间的变化则较小：从政策前到政策后，在参加调查前就已知道该政策的受访者比例从89%上升到98%，受访者所感知的因不使用塑料袋而带来的不便程度也略有增加。然而，平均而言，受访者对该政策减少环境污染目标的理解，对该政策的支持态度，所感知的政策效果，以及所感知的中国环境问题的严重程度均在政策后有所下降。政策前受访者对塑料袋难降解性的理解平均而言要强于政策后面临0.5元实际价格的受访者，却要弱于政策后面临0.3元实际价格的受访者。

受访者个人及其家庭的社会经济特征均值如下：受访者年龄平均为41岁，44%为男性；近20%的受访者持农业户口，近10%的受访者从事销售工作或自营职业，从事这类工作的消费者可能会因其工作所需的购物而比其他消费者消费更多塑料袋；受访者的受教育年限和月收入的均值分别为12.7年和2200元，而其平均家庭规模接近3人。许多社会经济变量的均值在政策前后的样本中存在差异。确切地说，从政策前到政策后，受访者中的男性比例，从事销售工作或自营职业的比例，农业户口比例，受教育年限和家庭规模的均值略有增加，而年龄则略有下降。政策前受访者的平均月收入高于政策后面临0.3元实际价格的受访者，而低于政策后面临0.5元实际价格的受访者。如表5最后三列的双侧 t 检验、双侧比例(proportion)检验和双侧Wilcoxon

表5 本研究使用的分析所使用变量的定义和统计描述

变量名称	变量描述	政策前		政策后		全部受访者样本 [4]	列[1]和列[2]数据的统计检验的P值	列[1]和列[3]数据的统计检验的P值	列[2]和列[3]数据的统计检验的P值
		0.3元和0.5元假设价格样本 [1]	0.3元实际价格样本 [2]	0.5元实际价格样本 [3]	0.5元实际价格样本 [3]				
实际价格	=1 如果受访者所面临的是实际价格；否则=0	0.000	1.000	1.000	1.000	0.608	0.000 ^b	0.000 ^b	1.000 ^b
收费塑料袋的比 例	=收费塑料袋占所消费的全部塑料袋的个数百分比 (%)	100.000	42.497	46.923	46.923	65.042	0.000 ^a	0.000 ^a	0.025 ^a
知道政策	=1 如果受访者在参与调查之前已经知道该限塑政策；否则=0	0.887	0.987	0.959	0.959	0.948	0.000 ^b	0.000 ^b	0.000 ^b
理解政策目标的程度	用1—5的标尺衡量受访者对政策目标的认识：=1为帮助商家变相牟利；=5为减少环境污染	4.720	4.173	4.053	4.053	4.387	0.000 ^c	0.000 ^c	0.426 ^c
对政策的态度	用1—5的标尺衡量受访者对政策的支持程度：=1为完全不支持；=5为完全支持	4.459	4.049	4.142	4.142	4.210	0.000 ^c	0.000 ^c	0.009 ^c
不用塑料袋的不便程度	用1—5的标尺衡量受访者所感知的由于不用塑料袋而带来的不便程度：=1为没有不方便；=5为非常不方便	2.746	2.974	3.124	3.124	2.885	0.000 ^c	0.000 ^c	0.039 ^c
政策有效性	用1—5的标尺衡量受访者所感知的政策有效性：=1为完全没用；=5为非常有效	3.554	3.250	3.337	3.337	3.369	0.000 ^c	0.001 ^c	0.089 ^c

(续表)

变量名称	变量描述	政策前			政策后			全部受访者样本 [4]	列[1]和列[2]数据的统计检验的P值	列[1]和列[3]数据的统计检验的P值	列[2]和列[3]数据的统计检验的P值
		0.3元和0.5元假设价格样本 [1]	0.3元实际价格样本 [2]	0.5元实际价格样本 [3]	0.3元实际价格样本 [3]						
理解塑料袋的难降解性	=1 如果受访者了解塑料袋难降解的特性是其对环境造成危害的最主要原因;否则=0	0.527	0.544	0.467	0.537	0.403 ^b	0.057 ^b	0.011 ^b			
环境问题严重性	用1—5的标尺衡量受访者所感知的中国面临的环境问题的严重性;=1为完全没问题;=5为问题非常严重	4.367	4.217	4.213	4.276	0.000 ^c	0.013 ^c	0.345 ^c			
年龄	= 受访者年龄(年)	42.841	40.528	38.482	41.435	0.001 ^a	0.000 ^a	0.040 ^a			
男性	=1 如果受访者为男性;否则=0	0.420	0.455	0.444	0.442	0.076 ^b	0.448 ^b	0.695 ^b			
农业户口	=1 如果受访者户口类型为农业户口;否则=0	0.181	0.197	0.225	0.191	0.307 ^b	0.075 ^b	0.246 ^b			
生意人	=1 如果受访者是从事销售工作或自营职业;否则=0	0.094	0.098	0.112	0.096	0.766 ^b	0.334 ^b	0.419 ^b			
受教育年限	= 受访者上学的年数(年)	12.371	12.844	13.003	12.658	0.000 ^a	0.002 ^a	0.402 ^a			
月收入	= 受访者的净月收入除以1 000(千元)	2.183	2.140	2.563	2.157	0.507 ^a	0.001 ^a	0.000 ^a			
家庭规模	= 受访者家庭中生活在一起的家庭成员数(人)	2.871	2.957	3.092	2.923	0.121 ^a	0.011 ^a	0.128 ^a			
观察值数		1 028	1 594	338	2 622	—	—	—			

注:^a表示该结果来自双侧t检验;^b表示该结果来自双侧比例(propotion)检验;^c表示该结果来自双侧Wilcoxon秩和(rank-sum)检验。

秩和 (rank-sum) 检验¹²的结果显示,许多变量的均值在列 [1] 和 [2] 的子样本之间以及在列 [1] 和 [3] 的子样本之间存在显著差异。由本研究大样本特征所产生的较小的变量标准误可能促成了这些差异的统计显著。然而,大多数社会经济变量的均值在列 [2] 和 [3] 的子样本之间基本相同,表明在政策后消费者所面临的塑料袋价格差异并非通过其自选择形成。

四、检验结果

(一) 统计检验

表6报告了在0.3元和0.5元价格下分别对本文第二部分所给出的研究假设进行相关统计检验的结果。在两价格下,我们分别采用双侧 t 检验和双侧Wilcoxon秩和检验对预测消费量与实际消费量的均值和分布相同的原假设进行检验。此外,我们还采用双侧比例检验对预测和实际完全停止消费塑料袋的受访者比例相同的原假设进行检验。

表6 统计检验结果

原假设1 (价格为0.3元)	$Q_{p=0.3}^A = Q_{p=0.3}^P$
塑料袋消费量均值的差异	0.034
t 检验 (P 值)	0.490
Wilcoxon秩和检验 (P 值)	0.072
观察值数: $Q_{p=0.3}^A / Q_{p=0.3}^P$	1594/1028
未消费塑料袋受访者所占比例差异	13.13%
比例检验 (P 值)	0.019
观察值数: $Q_{p=0.3=0}^A / Q_{p=0.3=0}^P$	659/290
原假设2 (价格为0.5元)	$Q_{p=0.5}^A = Q_{p=0.5}^P$
塑料袋消费量均值的差异	-0.103
t 检验 (P 值)	0.195
Wilcoxon秩和检验 (P 值)	0.648
观察值数: $Q_{p=0.5}^A / Q_{p=0.5}^P$	338/1028
未消费塑料袋受访者所占比例差异	12.92%
比例检验 (P 值)	0.000
观察值数: $Q_{p=0.5=0}^A / Q_{p=0.5=0}^P$	164/366

注: Q_p^P 表示在假设价格 p 下预测的塑料袋消费量; Q_p^A 表示在实际价格 p 下实际的塑料袋消费量。

统计检验结果表明在0.3元和0.5元价格下,所有检验均不能在5%或更高的显著水平上拒绝塑料袋的预测消费量与实际消费量的均值和分布无差异

¹² Wilcoxon秩和检验被用来对具有有序列型 (ranked) 数据结构的变量进行检验。

的原假设。¹³也就是说,在政策变化前运用陈述性偏好法让消费者和政策变化后的塑料袋消费量所进行预测的结果准确反应其实际消费量。完全停止消费新塑料袋的受访者所占比例的预测值和实际值却在两种价格下均存在显著差异。即在政策前调查中预测完全停止消费塑料袋的受访者比例显著低于政策后调查所记录的实际完全停止消费塑料袋的受访者所占比例。在考虑上述统计检验结果时读者应注意,由于上述检验并未控制诸如政策执行力度等可能对预测准确性产生影响的其他因素,因而结果还不具备十足的说服力。为了在剔除其他因素的可能影响后,对消费者预测和实际的塑料袋消费量之间的差异进行更严谨的检验,我们需要构建经济计量模型来进行多元检验。

(二) 多元检验

多元检验是在采用计量模型控制其他潜在影响因素的基础上对消费者塑料袋的预测消费量与实际消费量之间的潜在差异进行检验。虚拟变量“实际价格”的系数是显示两消费量差异的指针。表7报告了在0.3元和0.5元价格下分别对原假设进行多元检验的结果。¹⁴为简便起见,表7中略去报告模型所引入进行控制调查开展的商店和开展时间变量的系数。

表7 两种价格下对陈述性偏好法外部有效性检验的普通最小二乘模型回归结果

因变量	单次购物塑料袋消费个数	
	$p=0.3$ 元	$p=0.5$ 元
模型设定	[1] OLS 模型 I	[2] OLS 模型 II
	估计系数	估计系数
实际价格	-0.103 (0.073)	-0.121 (0.123)
收费塑料袋的比例	0.001 (0.001)	-0.003 (0.002)
知道政策	0.092 (0.106)	0.056 (0.110)
理解政策目标的程度	0.039 (0.029)	0.018 (0.042)
对政策的态度	-0.065 (0.035)*	-0.100 (0.049)**
不用塑料袋的不便程度	0.084 (0.019)***	0.056 (0.025)**
政策有效性	-0.082 (0.025)***	-0.030 (0.033)
理解塑料袋的难降解性	0.088 (0.048)*	0.072 (0.064)
环境问题严重性	-0.021 (0.030)	-0.051 (0.039)
年龄	0.009 (0.009)	0.016 (0.012)
年龄平方	-0.0002 (0.0001)**	-0.0002 (0.0001)**
男性	-0.129 (0.193)	-0.371 (0.275)
男性×对政策态度	0.053 (0.045)	0.085 (0.061)

¹³ 与 Shogren *et al.* (1999) 的发现一致,价格越高,预测值与实际值之间的差距越小。换言之,受访者在较高价格下对自己消费行为的预测会比在较低价格下更准确。产生这一现象的原因:一是较高价格可能刺激受访者更认真地进行考虑后才做出预测,二是0.5元作为首先给出的让受访者进行预测的价格,可能会使他们比对待后来出现的价格更认真地进行考虑和预测(Bateman *et al.*, 2004)。

¹⁴ 运用统计软件 Stata 的回归后命令计算模型估计量的方差膨胀因子(variance inflation factor)表明,模型不存在多重共线性问题。

(续表)

因变量	单次购物塑料袋消费个数	
	$p=0.3$ 元	$p=0.5$ 元
塑料袋价格	[1] OLS 模型 I	[2] OLS 模型 II
模型设定	估计系数	估计系数
农业户口	0.029 (0.068)	0.030 (0.091)
生意人	0.099 (0.084)	0.043 (0.112)
受教育年限	0.034 (0.032)	0.034 (0.041)
受教育年限平方	-0.002 (0.001)	-0.001 (0.002)
月收入	0.142 (0.040)***	0.140 (0.052)***
月收入平方	-0.012 (0.004)**	-0.010 (0.006)*
家庭规模	0.031 (0.017)*	0.009 (0.023)
调查开展时间虚拟变量	已控制	已控制
调查开展商店虚拟变量	已控制	已控制
观察值数	2622	1366
拟合优度(R-square)	0.156	0.097
Prob>F	0.000	0.000

注:(1)括号中所显示为标准误;(2)*表示在10%上显著;**表示在5%上显著;***表示在1%上显著。

多元检验结果显示,在控制其他因素的影响后,两价格下的模型估计结果均显示“实际价格”变量的系数为负,表明塑料袋的实际消费量在政策前被高估的趋势,但模型的双侧 t 检验显示该系数在两价格的模型中均未显著异于零。也就是说,采用多元检验我们仍不能在5%或更高的显著水平上拒绝塑料袋的预测消费量与实际消费量不存在差异的原假设。该结果表明,消费者在政策后实际支付塑料袋价格对其消费量的影响与政策前在陈述性偏好法假设情景下支付塑料袋价格对其消费量的影响的方向和程度均一致。换言之,通过陈述性偏好法在政策前获得的0.3元和0.5元假设价格下塑料袋消费的条件价值行为均准确预测了政策后同样实际价格下消费者的实际消费行为。

在两价格下进行的模型中,许多其他变量都表现出对塑料袋消费量具有显著影响。由于本文的主要兴趣在于检验陈述性偏好法的外部有效性,模型中引入其他变量是为了在控制它们的潜在影响基础上进行更精确的检验,因此本文将不再提供对其他变量影响的详细讨论。¹⁵此外,我们还尝试在模型中分别引入各种社会经济变量和控制变量分别与“实际价格”虚拟变量相交的各种交互变量,但未发现存在任一消费者子群体在政策后的实际消费行为能够在政策前更好或更坏地的由陈述性偏好法预测。¹⁶

¹⁵ 关于其他影响因素和该限塑政策对消费者塑料袋消费和使用行为的详细分析可见 He(2012)。

¹⁶ 我们采用了似然率(likelihood ratio)检验来对模型中所引入交互变量的共同显著性进行检验,发现不能拒绝这些交互变量共同不显著的原假设,故在本文所用的多元检验模型中未引入交互变量。

五、结 论

运用陈述性偏好法预测经济人行为需满足一个最基本的前提,即通过该方法在假设情景下估计出的条件价值行为能够准确对应相同人群在实际情况下的实际行为。本文通过将一项全国性环境政策法规的执行作为自然实验,在真实的市场环境中开展大规模调查,运用陈述性偏好法预测消费者对该政策进行规制的私人物品在政策变化后的消费行为,再以政策后调查所记录的实际消费行为为标准,对预测的准确性进行检验。研究结果表明,对普通消费者采用陈述性偏好法能够准确预测其对所熟悉的私人物品在相关公共政策变化后的实际消费行为。因此,本文的研究发现为陈述性偏好法能够准确预测公共政策的效果提供了积极的证据支持。

据我们所知,先前几乎所有讨论陈述性偏好法估计外部有效性的研究均在西方发达国家进行,而发展中国家的社会经济背景可能会在一定程度上影响这一方法所估计结果的外部有效性。在本文所研究的案例中,相较于发达国家的消费者,中国消费者总体上较少关心塑料袋对环境的危害¹⁷可能减少了他们陈述自己偏好时因“顾及脸面”(warm glow)而产生的影响,也就是说,中国消费者在政策前回答陈述性偏好法的假设问题时可能存在较小的希望让自己表现得环境友好的动机,从而有助于他们对自己塑料袋消费量进行较准确的预测。

本研究还总结出两点重要经验:第一,与 List and Gallet (2001) 的研究结果一致,受访者对塑料袋的熟悉以及塑料袋的私人物品性质可能有助于运用陈述性偏好法对其消费行为进行较准确预测;第二,鉴于 0.5 元是政策前根据其经营成本预测的在政策后最可能的塑料袋价格,陈述性偏好法对所有受访者均是从该价格开始询问,如果能更好地考虑到各商店会由于激烈竞争的市场环境而在政策后调低塑料袋价格,在政策前调查所运用的陈述性偏好法中,随机变换各种价格的出现顺序则应能更好地考虑可能存在的“顺序效应”(order effect)。最后,本研究结果能否适用于预测其他类型的政策和物品,则需通过未来开展更进一步的研究来进行回答。

¹⁷ 例如在我们的调查中仅有约一半的受访者了解塑料袋的难降解性是其对环境造成巨大危害的根本原因。

参 考 文 献

- [1] Bateman, I., M. Cole, P. Cooper, S. Georgiou, D. Hadley, and G. Poe, "On Visible Choice Sets and Scope Sensitivity", *Journal of Environmental Economics and Management*, 2004, 47 (1), 71—93.
- [2] Burtless, G., "The Case for Randomized Field Trials in Economic and Policy Research", *Journal of Economic Perspectives*, 1995, 9 (2), 63—84.
- [3] Carson, R., N. Flores, K. Martin, and J. Wright, "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods", *Land Economics*, 1996, 72 (1), 80—99.
- [4] Chang, J., J. Lusk, and F. Norwood, "How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior?" *American Journal Agricultural Economics*, 2009, 91 (2), 518—534.
- [5] Convery, F., S. McDonnell, and S. Ferreira, "The Most Popular Tax in Europe? Lessons from the Irish Plastic Bags Levy", *Environmental and Resource Economics*, 2007, 38 (1), 1—11.
- [6] Danish Environmental Protection Agency, "Waste in Denmark", Ministry of Environment and Energy, Copenhagen, Denmark, 1999.
- [7] Dikgang J., A. Leiman, and M. Visser, "Elasticity of Demand, Price and Time: Lessons from South Africa's Plastic-bag Levy", *Applied Economic*, 2012, 44(26), 3339—3342.
- [8] Hasson, R., A. Leiman, and M. Visser, "The Economics of Plastic Bag Legislation in South Africa", *South African Journal of Economics*, 2007, 75 (1), 66—83.
- [9] He, H., "Can Stated Preference Methods Accurately Predict Responses to Environmental Policies? The Case of a Plastic Bag Regulation in China", University of Gothenburg, Working paper No. 453, 2010.
- [10] He, H., "Effects of Environmental Policy on Consumption-Lessons from the Chinese Plastic Bag Regulation", *Environment and Development Economics*, 2012, Forthcoming.
- [11] 何浩然、陈安来, "中国限制塑料袋使用的政策效果及国际经验借鉴", 《中国人口、资源与环境》, 2010年第11期, 第167—174页。
- [12] Lampi E., and M. Orth, "Who Visits the Museums? A Comparison between Stated Preferences and Observed Effects of Entrance Fees", *Kyklos*, 2009, 62 (1), 85—102.
- [13] Levitt, S., and J. List, "What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?", *Journal of Economic Perspectives*, 2007, 21 (2), 153—174.
- [14] List, J., and C. Gallet, "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics*, 2001, 20 (3), 241—254.
- [15] Lusk, J., J. Pruitt, and B. Norwood, "External Validity of a Framed Field Experiment", *Economics Letters*, 2006, 93 (2), 285—290.
- [16] Mitchell, R., and R. Carson, "Using Surveys to Value Public Goods: The Contingent Valuation Method", Washington DC: Resources for the Future, 1989.
- [17] Shogren, J., J. Fox, D. Hayes, and J. Roosen, "Observed Choices for Food Safety in Retail, Survey, and Auction Markets", *American Journal of Agricultural Economics*, 1999, 81 (5), 1192—1199.
- [18] 中华人民共和国国家发展和改革委员会, 《国家发展改革委公告》, 2008年第33号, 第1—2页。

[19] 中华人民共和国商务部、中华人民共和国国家发展和改革委员会、中华人民共和国国家工商行政管理总局,《商品零售场所塑料购物袋有偿使用管理办法》,2008年第8号,第1—3页。

Can Impacts of Public Policies Be Accurately Predicted? Natural Experiment Evidence from the Plastic Bag Regulation in China

HAORAN HE

(Beijing Normal University & University of Gothenburg)

Abstract Chinese plastic bag regulation provides a natural experiment to investigate prediction validity of stated preference (SP) estimates on policy impacts. We utilized an SP survey to elicit consumers' contingent bag consumption in a series of hypothetical pricing scenarios before the regulation. In the regulation regime, we conducted another survey to record actual bag consumption. Results indicated that SP techniques predict the impacts of a policy change on consumers' behavior reasonably well.

JEL Classification C93, D12, Q53