

经济学实验结果外部效度排序探索研究

黄湛冰 肖尔特*

摘要: 经济学实验在政策研究方面具有独特优点,但针对同一问题的不同实验设计,其结论的外部效度并不相同,明确现有实验研究结果的外部效度差异可以为解决实际问题提供更精确的政策制定参考,并改进相应实验设计。目前针对具体问题的经济学实验结果外部效度量化排序方法尚属空白,本文基于 Harrison 和 List(2004)以及 Levitt 和 List(2007)的思想提出一个对经济学实验结果外部效度进行排序的两标准判断法,两个标准分别强调实验设计类型及实验处理中的个体行为因素,并通过对嵌套社会困境环境下的公共品自愿贡献实验结果进行外部效度排序说明这一方法的操作。

关键词: 经济学实验外部效度 排序方法 政策制定参考 嵌套社会困境 公共品实验

一、引言

经济学实验在政策研究方面具有一些其他研究方法所没有的独特优点^①,Roth 用“在王子耳边低语”(Roth,1988;Kagel and Roth,1995)比喻经济学实验研究结果在解决现实问题上的参考价值。然而,来自实验研究的“低语”通常具有不同的外部适用性,明确现有实验研究结果的外部效度差异可为解决实际问题提供更精确的政策制定参考,并为改进相应实验设计提供指引,如何判断一个具体问题下的多个实验研究结果的适用性差异?本文提出一个经济学实验结果外部效度排序方法,并通过对嵌套社会困境环境下的公共品实验结果外部效度进行排序说明其具体运用。

对经济学实验结果外部效度排序可为解决实际问题提供更精确的政策制定参考。从当今公共管理实践看,处在网络化环境中的公共管理者要处理越来越多复杂的、难以解决的奇特问题(Moore,2011),这类公共管理问题超越了单个组织的能力,并且相互联系,解决这类实际问题迫切需要得到相关理论的指导和建议;与此同时,大量的经济学实验研究成果揭示出不同因素间的内在关系,但数量日益庞大的实验结果及其不同的适用范围使政策制定者难以适应^②。在此情形下,对具体问题的经济学实验结果外部效度排序能够比较这些成果的外部适用性差异,帮助政策制定者对实际中能否采取与实验处理相似的做法做出更精确的判断和取舍,增加现有实验研究结果在解决现实问题上的参考价值。

本研究在实验设计改进上具有一定的理论价值。实验是否能反映现实的并行问题多年前就受到关注,通过对现有实验研究结果外部效度排序,可以探索具体问题上的实验研究结果和实际应用之间的并行关系,获得具体问题下的现有实验研究跟现实需求之间的缺口信息,为改进现有实验设计或开发新的实验设计提供知识。

* 黄湛冰,陕西师范大学国际商学院,邮政编码:710062,电子信箱:hzb10@163.com;肖尔特,卡耐基梅隆大学社会和决策科学系,邮政编码:15213。

感谢教育部人文社科规划基金项目“提高跨地区合作的嵌套社会困境公共品实验研究”(12YJA630041)的资助。特别感谢匿名审稿专家对本文提出的改进建议,当然文责自负。

①经济学实验的优点:较高程度的控制、可探索反事实、可复制、可预测、低成本等,这些优点使以实验为手段的研究在多个涉及公共政策的领域呈现不断增加的趋势(Alm,2011)。

②社会实验可在一定程度上对政策影响做出评价,但面临成本、时机、样本等组织实施问题,并难以在一次实验中针对性地探索多个因素的作用机理和实际影响。

目前有两种对经济学实验外部效度进行判断的理论性研究与本文相关:一种是针对实验类型,认为不同的实验设计类型具有不同的外部效度(Harrison and List, 2004);另一种是考察实验设计中是否包含物质财富之外的因素,认为同一种实验类型下的实验结果外推需要考虑非财富因素的作用(Levitt and List 2007)。但还没有出现一个能够对同一问题下已有实验结果外部效度进行排序的具体方法,这一缺失削弱了经济学实验在政策制定参考方面的作用发挥。本文基于Harrison和List(2004)以及Levitt和List(2007)的思想提出一个经济学实验研究外部效度量化排序的具体方法,填补这方面空缺。

以下先提出外部效度排序的标准及方法,再将此方法应用于嵌套社会困境公共品实验结果的外部效度排序以说明其应用及对实践的参考,最后提出嵌套社会困境公共品实验及研究可改进之处。

二、实验结果外部效度排序的标准选择及排序方法

Harrison和List(2004)以及Levitt和List(2007)分别从实验类型和个体的行为两个方向对实验的外部效度问题进行了探讨,本文分别将这两个方向作为对实验结果进行排序的第一标准和第二标准,两个标准与实验结果外部效度排序的关系如图1所示。

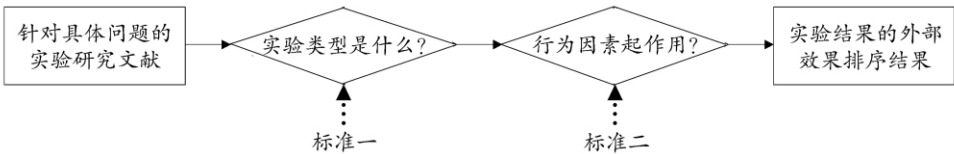


图1 实验结果外部效度排序与两个标准的关系

(一) 第一个标准

标准一:不同类型经济学实验研究结果具有不同的外部效度。

不同类型的实验设计有比较明显的外部效度差异。实验室实验(Conventional lab experiment,记为Lab)的特点是以学生作为被试,实验环境由一套抽象框架和强加的规则构成;伪实地实验(Artefactual field experiment,记为AFE,也称人工实地实验)的特点是除了被试对象采取非学生之外,其他方面与实验室实验相同;框架实地实验(Framed field experiment,记为FFE)的特点是在具体物品、任务、信息、资金、时间范围等方面采取实际中的物品和内容,其他方面与伪实地实验相同;自然实地实验(Natural field experiment,记为NFE)的特点是环境变成被试在现实中承担这些任务的场所,被试不知道自己在实验中,其他方面与框架实地实验相同;自然实验(Natural experiment,记为NE)的特点不受实验者控制(Levitt and List 2009)。

标准一的主要作用是通过对相关实验研究的实验设计对类型进行划分得到其实验结果外部效度的初步排序,如图2所示,外部效度自左往右增大。运用这一标准,首先可将所有的实验研究划分为三大类:实验室实验、实地实验、自然实验,其中实验室实验和自然实验分别处在效度坐标最外端,各自具有最强的内部效度和外部效度,位于中间的是实地实验;其次将实地实验再细分为伪实地实验、框架实地实验、自然实地实验三种(Harrison and List, 2004),实地实验处在完全受控和完全不受控之间,兼具实验室实验和自然实验的优点。



资料来源: Levitt 和 List(2009)。

图2 不同类型经济学实验的外部效度特点比较

将上述不同实验类型外部效度进行简单量化赋予序次,如表1所示,在表1“排序”行中,数字1、2、3、4、5代表外部效度次序,其中1代表该实验类型具有最高的外部效度,5代表最低外部效度。

表1 标准一规定的排序次序

实验类型	实验室实验(LAB)	实地实验(NE)			自然实验
		AFE	FFE	NFE	
排序	5	4	3	2	1

(二) 第二个标准

标准二: 实验环境中行为经济因素是否起作用以及作用方向、作用大小影响实验研究结果的外部效度。

1. 行为经济因素的重要性

在新古典经济学中人的行为基于以下假设: 个体是理性、结果导向、自我控制、完全自利、自我中心主体, 作为分析个体行为的起点这些假设很有用, 但大量的证据表明这些假设对个体行为的描述并不准确, 现实中的个体通常不具备标准理论所假设的那些特质, 由此催生了行为经济学, 行为经济学将其他社会科学(特别是心理学)的研究方法和证据运用到对个体和群体的决策分析(Angner and Loewenstein 2010) , 行为经济学把“偏离(departures)”新古典传统假设作为现实中对人的行为观察的起点, 在过去 20 年间借助分析经济现象中的行为经济因素, 行为经济学获得了大量新颖的理论认识和创新见解, 其发展改变了经济学家对个体行为的传统看法, 并进而影响到相应的政策制定。

在经济学、管理学的发展史上, 行为因素始终与这些学科的发展相随: 从亚当·斯密(Adam Smith)的道德情操论、埃奇沃思(Francis Edgeworth)的数学心理、费雪(Irving Fisher)的货币幻觉、凯恩斯(Maynard Keynes)的通论、西蒙(Herbert Simon)的管理行为、卡通纳(George Katona)的经济行为心理分析以及西多夫斯基(Tibor Scitovsky)的人类获得满足的心理等经典经济学管理学思想或理论都结合或者强调了决策过程中的行为因素(Alm 2011) 。

2. 行为公共经济学中的行为经济因素

行为经济学应用到公共经济学被称为行为公共经济学(Behavioral public economics) , 行为公共经济学运用行为经济学、心理学和神经科学的观点分析公共政策设计, 被认为在关键的政策上极具应用潜力, 不仅能够提出新颖和深刻的见解, 而且是产生开创性新方法的基础(Bernheim et al. 2007) 。

近来的研究探索了许多行为经济因素: 利他、不平等厌恶、互惠等等, 实验结果表明个体的行为会被不同的行为经济因素所驱动, 因此基于行为因素的大量实验处理(如沟通、亲社会惩罚、反社会惩罚、排斥放逐、身份认同等) 均可不同程度地影响公共政策。

3. 受行为因素影响的实验结果外部效度排序判断标准

如果认为标准一是对实验类型进行的划分, 那么标准二是在标准一划分的基础上, 对同一实验类型下的研究进行的再划分。

标准二源自 Levitt 和 List(2007) 的观点, Levitt 和 List(2007) 提出一个如式(1) 所示的模型:

$$U_i(a, p, n, s) = M_i(a, p, n, s) + W_i(a, p) \quad (1)$$

式(1) 表示在实验室中人的行为由财富与非财富两类因素共同决定。其中 U 表示个体 i 的效用, 它是行动 a (action) 、实验报酬刺激大小 v (stake) 、社会规范 n (norms) 、观察效应 s (scrutiny) 等变量的函数, U 的构成由两部分表示, 一个是非财富因素 M (在 Levitt 和 List(2007) 中用到的词是道德, 本文用更普遍的行为因素替代) , μ 代表伴随行动的行为成本或收益, 另一部分是财富 W , 代表个体的物质财富。个人行动对他人财富影响的外部性越大、社会规范 n 越强、观察效应 s 越大, 行为因素成本 M 越高; 行动 a 、实验刺激 v 越大, 财富 W 越高。根据 Levitt 和 List(2007) 的模型, 在行为因素不起作用的场合, 实验室结果可以直接运用到实际当中去, 特别在实验室环境反映了现实问题重要方面的时候, 此时的实验结果是现实行为很好的比照。本文对 Levitt 和 List(2007) 的公式进行延伸, 用 V 表示对应于式(1) 效用 U 的外部效度, 可区分出以下四种不同的效度: $V(M_i^+(a, v, n, s), W_i(a, v))$, 指行为因素和财富因素一致, 行为因素强化了财富因素; $V(M_i^+(a, p, n, s), W_i(a, p))$ 指没有行为因素作用, 只有财富因素起作用的场合(这种情形正是上述 Levitt 和 List(2007) 表述的状况) ; $V(M_i^-(a, p, n, s), W_i(a, p))$ 指行为因素起到与财富因素相反的作用, 但行为因素作用并不强烈; $V(M_i^{--}(a, p, n, s), W_i(a, p))$ 指行为因素起与财富因素相反的作用, 而且行为因素效果强烈。那么有以下外部效度大小比较判断:

$$V(M_i^+(a, p, n, s), W_i(a, p)) > V(M_i^-(a, p, n, s), W_i(a, p)) \quad (2)$$

$$V(M_i^-(a, p, n, s), W_i(a, p)) > V(M_i^{--}(a, p, n, s), W_i(a, p)) \quad (3)$$

$$V(M_i^-(a, p, n, s), W_i(a, p)) > V(M_i^{--}(a, p, n, s), W_i(a, p)) \quad (4)$$

本文将上述三式所包含的关于外部效度差异运用于比较排序, 同时将排序适用范围从 Levitt 和 List(2007) 的只针对实验室实验扩展到其他实验类型。具体排序如下: 首先判断实验环境中是否存在行为经济

因素,然后判断行为经济因素的作用与财富是冲突还是一致,最后判断在存在冲突的情形下,行为因素的影响是否强烈。这一排序标准可用表2来表示,最终的外部效度显示在排序数字1、2、3、4上,1代表最高的外部效度,4代表最低。

表2 标准二规定的排序次序

M 是否存在	存在			否
M 与 W 的关系	冲突		一致	-
M 的强度	强	弱	-	-
排序	4	3	1	2

(三) 两标准共同作用下的外部效度排序

上述两标准结合在一起可生成一系列有序的外部效度排序数字,具体步骤如下:

首先收集针对具体问题的所有实验研究文献,按照标准一对这些研究的实验设计类型进行划分,将实验设计分为实验室实验和非实验室实验,再把非实验室实验分为实地实验和自然实验两类,将实地实验再划分为三种细类(图3中的六角形实线框表示),进行以上步骤主要依据是表1所体现的标准一。

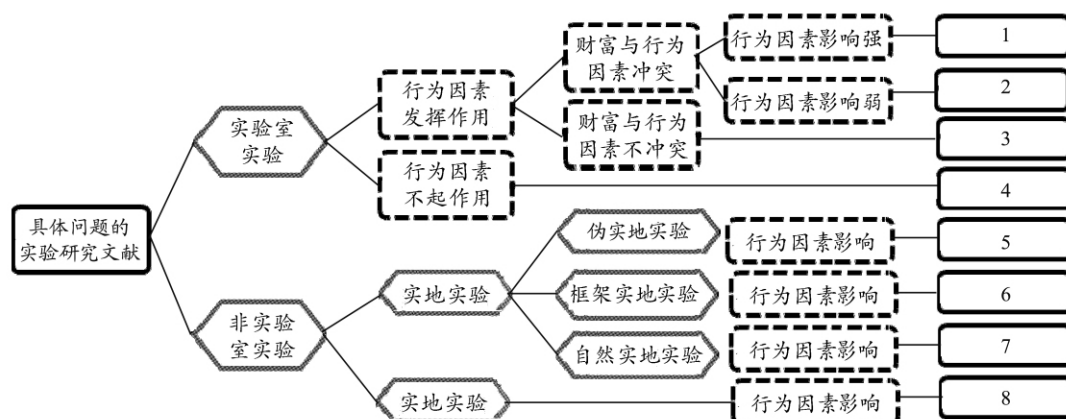


图3 符合两个标准的实验研究结果外部效度排序

接下来对每种类型下的实验运用标准二进行细分(图3中的虚线框表示,为简洁起见,图3只画出针对实验室实验类型的细分,其他类型可仿此进行),在标准二下的主要工作是区分行为因素是否起作用、作用方向、作用强度三方面,其比较排序参照表2进行,图3中“实地实验”、“自然实验”后的“行为因素影响”指与图3“实验室实验”后类似的排序处理。图3中的排序数字1代表最高的外部效度,8代表最低的外部效度(需要说明的是,类似“实验室实验”后面的处理每种类型在标准二下可排出4个外部效度次序,因此全部5种实验类型完整排序下来可以排列出20种外部效度,图3对除实验室实验之外的其他四种类型图示作了简化)。

三、实验研究结果外部效度排序举例: NSD 实验

下面用与跨地区合作问题相关的嵌套社会困境(Nested social dilemmas,简称NSD)实验研究为例,说明上述排序方法的具体应用。

随着区域间竞争加剧,作为应对竞争的重要策略,跨地区合作不仅能打破行政分割、降低交易成本,保证相关区域发展所需要的基本要素,并能抵御来自其他经济发达区域的极化作用,因此日益受到众多地方政府重视,但在实践中跨地区合作难问题一直没有得到有效解决,合作实践效果往往不理想甚至流于形式(周业安、宋紫峰,2008;陈瑞莲、谢宝剑,2009;张紧跟,2010),如何从理论研究上为解决这一实际问题提供政策制定参考?嵌套社会困境下的公共品研究是解决跨地区合作问题的一种理论途径。

(一) 跨地区合作问题的NSD公共品特点

跨地区合作问题具有社会困境和多公共品特点。社会组织中许多最基本的问题本质上是公共品问题(Ledyard,1995),跨地区合作可为相关地区带来共同利益,但每个地区同时还有搭便车的选择,因此这一问

题具有社会困境的特征^① ,由于合作问题强调各个地区的贡献 ,因此这种社会困境呈现出公共品特点^②。公共管理的最新发展趋势强调公民参与(Osborne ,2009) ,因此有必要进一步将每个地区的黑箱打开 ,把处于同一地区的利益相关者同时纳入考虑 ,进行这样的处理后可以发现跨地区的合作还具有明显的嵌套多公共品特征: 同时存在有益于各个地区当地福利的局部公共品和有益于整个区域的跨地区总体公共品 ,因此它是一种独特的、相对复杂的社会困境问题。

上述情形被称为嵌套社会困境(Nested social dilemmas ,简称 NSD) (Polzer ,et al. ,1999) ,即具有分层社会结构的社会困境。一般社会困境通常只是个体利益与群体利益间的冲突 ,决策者在个人收益和群体收益两个账户上进行权衡 ,决策只涉及群体收益一个公共品; 但现实中常见的实际情形是不仅个体是一个群体的成员 ,同时多个群体构成一个更大的集体 ,群体嵌套于集体中 ,决策者同时在个体、群体和集体三个账户中进行权衡 ,有时候甚至还包括多层嵌套社会结构关系。Polzer 等(1999) 指出这种群体公共品嵌套于集体公共品是一种很普遍的结构 ,不仅是跨地区合作问题 ,现实中的许多公共品问题如家族事务、组织关系、国际合作、环境治理、反恐安全等均具有嵌套社会困境的特点。

公共品实验是当前研究社会困境最常用的方法之一 ,目前国内还没有类似实验研究 ,笔者掌握到的国外 NSD 实验研究有 8 篇 ,这使得本文可在较小的工作量下对现有 NSD 实验研究成果外部效度进行排序。

(二) NSD 实验文献分析

文献分析是外部效度排序的基础 ,限于篇幅没有将其放入。

四、NSD 公共品实验外部效度排序判断

(一) NSD 研究涉及到的实验类型

对 NSD 实验研究文献首先按照标准一进行类型区分 ,结果如表 3 所示。

表 3 NSD 实验研究的实验类型区分

文献(按时间顺序)	实验类型	细分	被试身份
Wit 和 Kerr(1994 2002) ③	实验室	-	学生
Polzer 等(1999)	实验室	-	学生
Wachsman(2002)	实验室	-	学生
Blackwell 和 Mckee(2003)	实验室	-	学生
Fellner 和 Lunser(2008)	实验室	-	学生
Buchan 等(2009)	实地	伪实地实验	非学生
Rand 等(2009)	实地	框架实地实验	非学生
Dalton(2010)	实地	伪实地实验	非学生

标准一应用后 ,在其基础上再应用标准二 ,此时遇到标准一和标准二之间先后次序问题 ,在标准一和标准二之间的排序关系上 ,本文采取的是字典式排序偏好^④ ,将标准一放在标准二之前 ,就是说如果一个实验结果的外部效度按照标准一排序在前 ,那么无论按照标准二如何排序 ,按标准一排序在前的总是排序在前 ,只有当两个实验结果在标准一排序下相同时 ,才按照标准二排序。实验类型优先基于以下考虑:

1. 行为经济因素和环境间关系

行为经济学放松了传统经济学经济人假设 ,注重心理学规律、实验证据 ,从行为博弈的角度看行为经济学是在传统分析博弈的收益函数上加入行为修正部分 ,使之更符合现实(Camerer 2003) 。

许多研究表明行为效果在实验室环境下通常比较明显 ,而在现实环境中常减弱甚至消失(Levitt and List 2007) ,行为经济因素的作用基础是心理学规律(Camerer 2003) ,心理学认为心理学规律与环境密切相关(如情境心理学的人与情境互动论) 。

①社会困境是指个体最优的行动与大部分参与者都采取个体最优行动反而损害每个参与者利益的两难冲突情形。

②社会困境包括囚徒困境、公共地、公共品等不同形式 ,见 Polzer 等(1999) 。

③Wit 和 Kerr(1994) 是 Wit 和 Kerr(2002) 的工作论文版本。

④字典式排序偏好(Lexicographic preferences) 的定义为: 给定两个半序集 A 和 B ,如果在 $A \times B$ 笛卡尔积上 ,有 $(a , b) \leq (a' , b')$ 当且仅当 $a < a'$ (或者 $a = a'$ 且 $b \leq b'$) 时成立 ,就称为字典式排序偏好。

不同实验类型的环境存在差异,包括:实验参与者被观察的性质和程度(即霍桑效应)、实验场景和过程、实验中起作用的社会准则、时间长短、参加实验的个体类型、经验等多个方面(Levitt and List 2007)。

由于不同实验类型的实验环境存在差异,因此不同实验类型下产生的心理行为及行为经济因素强度大小不相同,在一定程度上可认为行为经济因素强度差异取决于实验类型间的环境差异。

2. 实验真实性

外部效度涉及不同实验类型的真实性,对经济学实验的“真实性”目前存在认识争议:Smith(1994, 2003)、Plott(1986)等认为实验室实验和实地实验一样具有真实性;Levitt和List(2007, 2009)认为实地实验比实验室实验更真实,更真实的实验与政策关系更紧密,结果也更容易外推到现实中去;Falk和Heckman(2009)认为不论什么形式的实验,当研究设计越反映问题本质,研究结果越容易外推(Bardsley, et al., 2009)。

为兼顾这些观点,在使用本方法时,可通过对标准二赋予不同权重:当倾向于Falk和Heckman(2009)观点时,标准二权重的大小由环境设计是否反映问题的本质决定,如果行为因素产生的研究环境能够反映问题本质就赋予标准二高权重;当倾向于Smith(1994, 2003)观点时,不同实验类型下的标准二权重可设定为相同。

本文采取List(2007, 2009)的观点^①,认为不同实验类型的真实性不同,同时不同实验环境下,行为经济因素作用的突出程度通常不一样,两者综合考虑,从实验室实验到实地实验再到自然实验,在真实性增加的同时,环境中的干扰因素通常也在增加,此时行为经济因素赖以产生的心理环境条件被改变,行为经济因素作用效果递减或消失,因此本文在评价行为经济因素外部效度时,将实验类型放在优先位置上。

(二) NSD 研究实验环境中的行为因素作用

将NSD实验研究文献分为不同实验类型后,运用标准二将那些用标准一无法区分开的实验研究进行再区分,重点分析实验设计中的处理差异,对行为因素作用 M 是否存在、作用方向、强弱大小进行判断比较。

M 是否存在及 M 的方向。对NSD中 M 是否存在及作用方向比较容易做出判断,因为社会困境就是一个涉及到 M 的问题,而且这种 M 的方向对财富的影响从个体的角度来看是冲突的,所以其作用方向相反。

对NSD实验中 M 的分析主要集中在区分其大小上,这一分析过程的核心在于分析实验设计特点,根据文献分析所揭示的实验设计及结论特点对相关实验外部效度进行区分。

1. 实验室实验

对实验室实验中 M 的大小判断具体步骤和结果如下:

首先将研究分为身份类属和环境变化两种实验设计,这样一来身份类属发生改变的研究有2个,环境变化的研究有3个。

其次分析两种实验设计下的更具体的实验处理差异。在身份类属都发生变化的情况下,禀赋的初始位置设计可以不同,根据这一禀赋初始位置差异可以区分相关研究。在环境变化的实验设计下,按照两个处理差异进行区分:一个是是否允许沟通,另一个是依据边际单位资本回报(Marginal Per Capita Return, MPCR)设计变化范围。

进行了以上区分后,再考虑行为因素,此时的主要工作是对已经区分开的研究进行两两比较,对实验的外部效度进行排序,以下以 M (设计差异)表示由于特定“设计差异”产生的行为因素 M 的影响作用,具体地有以下比较判断结果{ }中的文字为依据理由:

- (1) M (环境变化) < M (身份类属) { 外因的作用小于内因};
- (2) M (禀赋位置在个体帐户) < M (禀赋位置在非个体账户) { 框架效应}
- (3) M (无沟通) < M (沟通) { 沟通的作用已被一般公共品研究所确认}
- (4) M (MPCR变化范围小) < M (MPCR变化范围大) { 范围小的设计包含于大的设计中}

将以上判断综合起来得到表4所示结果。

^①不采取Smith(1994, 2003)观点主要考虑到迪昂-奎因整体论论点(Duhem-Quine thesis)(李醒民,1996)、不采取Falk和Heckman(2009)观点主要考虑到心理学中关于行为的跨情境一致性观点,“行为表现一方面要满足个体原发性的动机需要,即行为的一致性,或保持自我的统一性,但同时还要适应环境的要求,按照环境的要求调整自己的行为。在大多数情况下,行为的一致性往往会让位于适应环境的压力,至少在中国文化下如此。”(王登峰、崔红,2006)

表 4 NSD 实验室实验下的行为因素作用区分

实验室实验文献	实验设计特点	处理差异				M 的作用	排序
		身份类属	环境变化	禀赋与沟通	MPCR 变化		
Wit 和 Kerr (1994 2002)	随机装置实现类属的显著性	变	-	禀赋在个体账户	-	强	5
Polzer 等(1999)	初始禀赋分别放在群体及集体账户	变	-	禀赋在非个体账户	-	弱	4
Wachsman(2002)	边际报酬 MPCR 差异; 沟通差异	-	变	沟通	-	强	3
Blackwell 和 McKee (2003)	保持群体 MPCR 不变, 增加集体 MPCR	-	变	无沟通	大范围	弱	1
Fellner 和 Lunser (2008)	改变集体 MPCR, 增加集体效率	-	变	无沟通	小范围	次强	2

2. 实地实验

对实地实验中 M 的大小判断比较简单。因为框架实地实验(Rand et al. 2009)的外部效度大于伪实地实验,所以主要的工作在于对两个伪实地实验的外部效度比较,这一比较仍基于两个实验设计的差异:一个是关注群体间差异,另一个是关注个体间禀赋差异。越关注个体的实验设计越会在社会困境环境下产生行为成本,因此有以下判断结果: M (关注群体间禀赋差异) < M (关注个体间差异),如表 5 所示。

表 5 NSD 实地实验下的行为因素作用区分

实地实验文献	实验处理	身份类属	差异	M 的作用	赋值
Buchan 等(2009)	控制不同地域的全球化指数	变	关注群体间差异	弱	1
Dalton(2010)	控制收入愿望缺口	变	关注个体间禀赋差异	强	2

(三) NSD 研究实验结果外部效度排序结果

将表 4 和表 5 的结果结合在一起,可以得到全部 NSD 实验研究的外部效度排序结果,如表 6 所示。

表 6 NSD 实验研究的外部效度排序

文献	实验类型	细分	排序
Wit 和 Kerr(1994 2002)	实验室	-	8
Polzer 等(1999)	实验室	-	7
Wachsman(2002)	实验室	-	6
Fellner 和 Lunser(2008)	实验室	-	5
Blackwell 和 McKee(2003)	实验室	-	4
Dalton(2010)	实地	伪实地实验	3
Buchan 等(2009)	实地	伪实地实验	2
Rand 等(2009)	实地	框架实地实验	1

(四) 跨地区合作政策制定参考

以上排序在实验和实际应用参考之间建立起了一种联系,为决策制定者将 NSD 实验研究结果应用到现实问题的解决上提供了参考选择和取舍的依据,外部效度越高的实验研究结果越具有较高的实际参考价值和较高的采纳可能性,具体实验揭示的做法如表 7 所示(限于篇幅,本文不对解决 NSD 问题的具体政策制定参考做法进行展开)。

表 7 NSD 实验揭示的可参考做法及外部效度

作者	处理	外部效度排序
Wit 和 Kerr(1994 2002)	随机装置实现类属的显著性	8
Polzer 等(1999)	初始禀赋分别放在个人、群体及集体账户	7
Wachsman(2002)	边际报酬 MPCR 差异; 沟通差异	6
Fellner 和 Lunser(2008)	改变集体 MPCR, 增加集体效率	5
Blackwell 和 McKee(2003)	保持群体 MPCR 不变, 增加集体 MPCR	4
Dalton(2010)	收入愿望缺口	3
Buchan 等(2009)	不同地域的全球化指数	2
Rand 等(2009)	群体的身份发生动态变化	1

归纳 NSD 实验研究结果,得到一个有助于提高跨地区合作效果的政策制定思路:弱化群体身份类属,突出集体类属,促使个体认同集体目标。其主要依据是集体行动取决于个体的自我类属和意图(Wright, 2009),及集体行动是社会环境与个体间的相互作用结果(Levitt and List 2007)。在这一思路下,有两类具体方法,一类是改变个体身份,另一类是改变环境。两类方法中改变身份比间接改变环境具有更高的外部效度,因此也更具有操作可行性和政策制定参考价值,如图 4 所示。

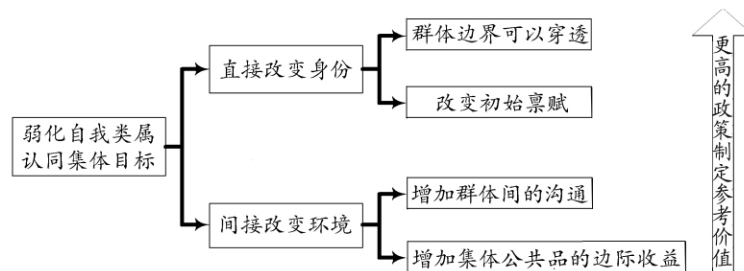


图 4 解决实际 NSD 问题的框架性思路

(五) 对今后实验设计及研究方向的指引

以上实验结果对今后实验设计及研究方向分别有以下启示:

1. 重视实地实验

将表 7 的外部效度排序结果绘制成趋势图,如图 5 所示,图中纵坐标表示外部效度大小,为符合效度高数字大的习惯,该值用表 7 第 3 列的排序数字减 9 取正获得。横坐标为年份,拟合曲线显示变化趋势。NSD 实验研究外部效度变化趋势表明外部效度高的 NSD 研究愈来愈受重视,预示今后的实验设计应多采取实地实验形式。

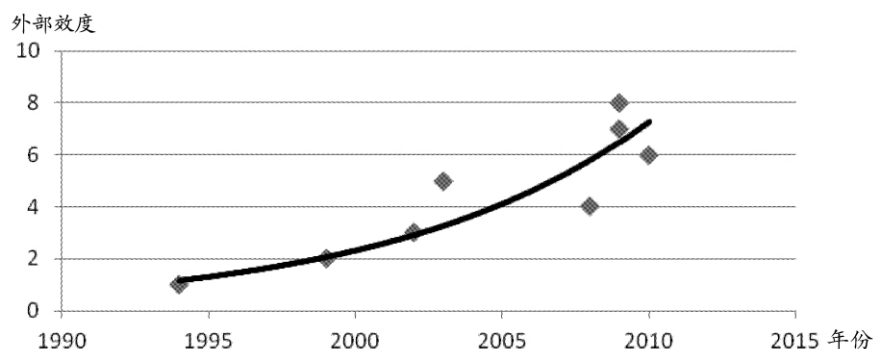


图 5 NSD 实验研究结果外部效度变化趋势

2. 探寻更多增加个体在集体公共品上贡献的方法

根据以上分析及多公共品目前研究进展,以下三个研究方向上的 NSD 实验设计值得重视:

(1) 相互监督和条件合作。相互监督和条件合作在单公共品环境下具有明显的作用,但尚未在 NSD 上进行研究。

(2) 差异化不同群体从集体公共品上获得的 MPCR。现有的 NSD 实验设计将集体公共品对每个群体产生的 MPCR 设定为相等,实际中更一般的情形是不同的群体从集体公共品上获得的 MPCR 存在差异。

(3) 绑定处理。在多公共品环境下会出现一种特别的处理:参与者在多个公共品上或者都不贡献,或者贡献相同(也可以是固定比例关系),这种限制称为绑定(Cornes 2010)。在实际中这种现象很常见,比如研究经费的配套关系等,但目前没有这类实验研究。

这些潜在的研究方向预示今后的研究需继续探寻更多增加个体在集体公共品上贡献的方法。

五、排序方法在新公共治理阶段的应用前景

公共管理理论经过行政管理(PA)和新公共管理(NPM)两个阶段后,目前进入到新公共治理(NPG)阶段(Osborne, 2009)。

在新公共治理阶段,人们意识到宏观、中观和微观之间的交互动态共同决定公共管理效果,纯粹由政府机构提供公共服务难以实施甚至不合需要,公民参与下的共同生产才是确保公共价值和利益的治理机制。

(Calabrò 2011) ,研究者和实践者认识到国家、地方政府、利益相关者(包括私人、企业、第三方组织等)之间形成互动关系,公众不再是公共服务的被动接受对象,而是成为具有主动参与公共事务管理的现实主体,因此公共价值的共同创造受到重视(Williams and Shearer 2011),公民参与变成公共管理不可或缺的要素。

如果将个体的行为反应纳入公共管理研究中,传统的数据收集方式将面临难以反映其中互动关系的困难,经济学实验便成为一种很有用处的研究手段,通过对研究对象进行良好的控制,研究者可以对新公共治理框架下的个体行为、群体行为、社会准则、多个主体间的互动行为等因素进行系列实验研究,收集到难以在现实中获取的多方面有用数据,进而以实验设计和研究内容为基础分析其外部效度排序,为具体的政策设计提供综合性的理论参考基础和决策依据。

基于以上认识,本文提出的两标准实验结果外部效度排序方法在新公共治理阶段的公共管理研究与实践中具有一定的应用前景,它可以对现有实验结果从一个综合的角度进行再认识,同时也能够对下一步的实验研究提出有益的指引。

六、结论与讨论

通过提出实验研究结果外部效度排序标准及对NSD实验结果进行外部效度排序,本文得到以下结论:

(1) 实验研究的实验设计类型和具体实验处理中包含的行为因素方向及强弱可作为两个判断相关研究外部效度大小排序的标准,分别起到实验类型外部效度排序和特定实验类型下外部效度进一步排序的作用,综合运用两个标准,能够得到实验研究结果外部效度的一个完整排序。

(2) 对NSD实验研究结果进行的排序过程说明,这一两标准综合排序方法具有可操作性,两标准排序方法可推广到其他问题实验研究结果外部效度排序上。

(3) 外部效度的排序结果在实验研究和实际应用之间建立起一种对应关系,通过这一排序结果向决策制定者提供了一个将实验研究结果外推到现实管理实践中去的参考选择依据。

(4) 该排序方法在新公共治理这一新的公共管理研究和实践趋势上有潜在的应用前景,新公共治理下,公共问题的解决需将公民参与和不同层面主体间互动放在一起研究,数据收集困难使实验受到青睐,实验结果与实践之间的联结需求使该方法具有价值。

本文方法有一定的适用范围,该方法适用于把实验作为模型来应用的情形^①,而在实验作为检测作用的场合,则应更注重内部效度。

参考文献:

1. 陈瑞莲、谢宝剑. 2009《回顾与前瞻: 改革开放30年中国主要区域政策》,《政治学研究》第1期。
2. 李醒民. 1996《迪昂》,台北三民书局。
3. 王登峰、崔红. 2006《行为的跨情境一致性及人格与行为的关系》,《心理学报》第4期。
4. 张紧跟. 2010《区域公共管理制度创新分析: 以珠江三角洲为例》,《政治学研究》第3期。
5. 周业安、宋紫峰. 2008《公共品的自愿供给机制: 一项实验研究》,《经济研究》第7期。
6. Alm, James. 2011. "Testing Behavioral Public Economics Theories in the Laboratory." Working Papers from Tulane University, Department of Economics, No. 1102.
7. Angner, E. and G. Loewenstein. 2010. "Behavioral Economics." In *Handbook of the Philosophy of Science*, Vol. 13: *Philosophy of*

^①以下说明参考了Bardsley等(2009)第2、4、5章。近来研究认为经济学实验有两类作用:一类是检测(test),为已有理论服务,先有演绎理论结果,然后设计实验作为检测工具,最后那些通过严格检测的实验结果填补理论结果和经验证据之间的空隙,其思路是建构主义或笛卡尔主义,特点是理论先行、实验效度主要考虑内部效度;另一类作用是模型,产生的原因在于演绎理论发展不充分,导致实验超越现有理论发展限制,在这种情形下中心是实验而不是演绎理论,研究者通过实验归纳在实验中发现的规律,形成新知识,其特点是实验先行。实验的模型作用反映了近来一些科学哲学工作者的观点,如Guala(1998)、Mäki(2005)、Morgan(2005)等,其中Mäki(2005)甚至提出“实验是模型,模型是实验”,这种观点与历史上许多发展事实相一致,比如历史上的内燃机发明,是基于物理模型而不是数学演绎模型发展起来。作为模型起作用的社会科学实验,需重视外部效度问题。在社会科学领域中,变量间的关系常比较复杂,实验者不断地超越理论发展,扩展实验的模型作用,当采取相同或相近的模型设计时,可认为其内部效度相近,此时外部效度成为实验总效度要考虑的重要部分;同时,社会科学领域中的实验,通常其外部效度受到批评和质疑,特别是在对政策进行指导时(Guala 2005),这类质疑增加了研究外部效度的价值。

Economics , ed. Uskali Mäki , 67 – 101. Amsterdam: Elsevier.

8. Bardsley , Nicholas , Robin Cubitt , Graham Loomes , Peter Moffatt , Chris Starmer , and Robert Sugden. 2009. *Experimental Economics: Rethinking the Rules*. Princeton: Princeton University Press.
9. Bernheim , J. D. , and Antonio Rangel. 2007. “Behavioral Public Economics: Welfare and Policy Analysis with Non Standard Decision Makers. ” In *Behavioral Economics and Its Applications* , ed. Diamond P. and H. Vartiainen. Princeton NJ: Princeton University Press.
10. Blackwell , J. , and M. McKee. 2003. “Only for My Own Neighborhood? Preferences and Voluntary Provision of Local and Global Public Goods. ” *Journal of Economic Behavior & Organization* , 52(1) : 115 – 131.
11. Buchan , N. R. , G. Grimalda , R. Wilson , M. Brewer , E. Fatas , and M. Foddy. 2009. “Globalization and Human Cooperation. ” *Proceedings of the National Academy of Sciences of the United States of America* , 106(11) : 4138 – 4143.
12. Calabrò , Andrea. 2011. *Governance Structures and Mechanisms in Public Service Organizations*. Heidelberg: Physica – Verlag HD.
13. Camerer , C. 2003. *Behavioral Game Theory*. Princeton NJ: Princeton University Press.
14. Cornes , Richard , and Jun – Ichi Itaya. 2010. “On the Private Provision of Two or More Public Goods. ” *Journal of Public Economic Theory* , 12(2) : 363 – 385.
15. Dalton , Patricio S. 2010. “Income Aspirations and Cooperation: Experimental Evidence. ” Working Paper of Tilburg University and Center , the Netherlands.
16. Falk , Armin and James J. Heckman. 2009. “Lab Experiments are a Major Source of Knowledge in the Social Sciences. ” *Science* , 326(5952) : 535 – 538.
17. Fellner , G. , and G. K. Lunser. 2008. “Cooperation in Local and Global Groups. ” Department of Economics , Vienna University of Economics & B. A , Working Paper Series No. 122.
18. Guala , F. 1998. “Experiments as Mediators in Non – laboratory Sciences. ” *Philosophica* , 62(2) : 57 – 75.
19. Guala , F. 2005. *The Methodology of Experimental Economics* 203 – 231. London: Cambridge University Press.
20. Harrison , Glenn W. , and John A. List. 2004. “Field Experiments. ” *Journal of Economic Literature* , 42(4) : 1009 – 1055.
21. Kagel , J. H. , and A. E. Roth. 1995. *The Handbook of Experimental Economics* 22. Princeton NJ: Princeton University Press.
22. Ledyard , O. 1995. “Public Goods: Some Experimental Results. ” In *Handbook of Experimental Economics* , ed. J. Kagel and A. Roth , 111 – 194. Princeton NJ: Princeton University Press.
23. Levitt , S. D. , and J. A. List. 2007. “Viewpoint: on the Generalizability of Lab Behavior to the Field. ” *Canadian Journal of Economics* , 40(2) : 347 – 370.
24. Levitt , S. D. , and J. A. List. 2009. “Field Experiments in Economics: The Past , the Present , and the Future. ” *European Economic Review* 53(1) : 1 – 18.
25. Mäki , Uskali. 2005. “Models are Experiments , Experiments are Models. ” *Journal of Economic Methodology* , 12(2) : 303 – 315.
26. Moore , Tim. 2011. “Wicked Problems , Rotten Outcomes and Clumsy Solutions: Children and Families in a Changing World. ” NIFTeY / CCCH Conference 2011 , Children’s Place on the Agenda: Past , Present and Future , Sydney.
27. Morgan , M. S. 2005. “Experiments Versus Models: New Phenomena , Inference and Surprise. ” *Journal of Economic Methodology* , 12(2) : 317 – 329.
28. Osborne , S. P. 2009. “Debate: Delivering Public Services: Are We Asking the Right Questions?” *Public Money and Management* , 29(1) : 5 – 7.
29. Plott , J. C. 1986. “Laboratory Experiments in Economics: The Implications of Posted – Price Institutions. ” *Science* , 232(4751) : 732 – 738.
30. Polzer , J. T. , K. J. Stewart , and J. L. Simmons. 1999. “A Social Categorization Explanation for Framing Effects in Nested Social Dilemmas. ” *Organizational Behavior and Human Decision Processes* , 79(2) : 154 – 178.
31. Rand , D. G. , T. Pfeiffer , A. Dreber , R. W. Sheketoff , N. C. Wernerfelt , and Y. Benkler. 2009. “Dynamic Remodeling of In – group Bias during the 2008 Presidential Election. ” *Proceedings of the National Academy of Sciences of the United States of America* , 106(15) : 6187 – 6191.
32. Roth , A. 1988. “Laboratory Experiments in Economics: A Methodological Overview. ” *The Economic Journal* , 98(393) : 974 – 1031.
33. Smith , V. L. 2003. “Constructivist and Ecological Rationality in Economics. ” *American Economic Review* , 93(3) : 465 – 508.
34. Smith , V. L. 1994. “Economics in the Laboratory. ” *Journal of Economic Perspectives* , 8(1) : 113 – 131.
35. Wachsman , Y. 2002. “The Effects of Group Interaction in a Public Goods Experiment with Two Exchanges. ” Department of Economics , University of Hawaii at Manoa , Working Paper 02 – 5.
36. Williams , I. , and H. Shearer. 2011. “Appraising Public Value: Past , Present and Futures. ” *Public Administration* , 89(4) : 1367 – 1384.

(下转第 74 页)

Blackwell.

13. Guo ,K. and Y. Yao. 2005. "Causes of Privatization in China. " *Economics of Transition* ,13(2) : 211 – 238.
14. Gupta ,N. ,J. Ham ,and J. Svejnar. 2008. "Priorities and Sequencing in Privatization: Evidence from Czech firm Panel Data. " *European Economic Review* 52(2) : 183 – 208 .
15. Hausman J. and D. McFadden. 1984. "Specification Tests for the Multinomial Logit Model. " *Econometrica* 52(5) : 1219 – 1240.
16. Megginson ,W. and J. Netter. 2001. "From State to Market: A Survey of Empirical Studies on Privatization. " *Journal of Economic Literature* 39(2) : 321 – 389.

State – owned Enterprises’ Efficiency and Their Choices for Exit: Empirical Evidence from Firm Level Study in Several Competitive Industries

Zhu Kepeng¹ and Liu Xiaoxuan²

(1: The School Economics ,Anhui University of Technology;

2: Institute of Economics ,Chinese Academy of Social Science)

Abstract: This paper is a study of determinants of State – owned Enterprises’ exit in several competitive industries of China in the period of 2000 – 2008. Measuring firms’ efficiency and defining different exiting choices in an unbalance penal data base , the paper builds series of Multinomial Logit Models for regression analysis. The results show that the efficiency is a very important determinant of SOEs’ choices for exit. Efficient SOEs are likely to choose privatization ,while less efficient SOEs choose bankruptcy or reorganization. However ,the importance of efficiency lowers significantly in late of the period of 2000 – 2008. Changes on efficiency behavior of SOEs reflect turns of government’ s targets.

Key Words: State – Owned Enterprises; Firm Efficiency; Firm Exit; Selection Bias; Privatization

JEL Classification: D21 ,J25

(责任编辑: 彭爽)

(上接第 29 页)

37. Wit , P. , and N. L. Kerr. 2002. "Me versus Just US versus All: Categorization and Cooperation in Nested Social Dilemmas. " *Journal of Personality and Social Psychology* 83(3) : 616 – 637.
38. Wright , S. C. 2009. "The Next Generation of Collective Action Research. " *Journal of Social Issues* 65(4) : 859 – 879.

Exploratory Research on External Validity Ranking Method of Experimental Results in Economics

Huang Zhanbing¹ and Xiao Erte²

(1. International Business School of Shaanxi Normal University;

2. Social and Decision Science Department of Carnegie Mellon University)

Abstract: Economics experiments have some special advantages over other analysis methods. The external validities of different economical experiments are usually different even in a same research direction. A policy making reference for real issues can be got by analyzing the experiments’ external validity differences. Although some studies have provided conceptual judgment and classification perspectives on external validity of economical experiments , concrete ranking method is scanty. This thesis puts forward to a method which can rank experiments’ external validities on the basis of Harrison & List(2004) and Levitt & List(2007) . The core of the method is consisted of two rules. They stress types and behavioral reaction in experiments design. Experiments on nested social dilemmas are analyzed as an applied example.

Key Words: External Validity of Economical Experiments; Ranking Order Rules; Policy Making Reference; Nested Social Dilemmas; Public Goods Experiments

JEL Classification: B41 ,C90 ,D71 ,H41 ,J18

(责任编辑: 彭爽)