

社会学定量分析中的内生性问题

测估社会互动的因果效应研究综述

社会
2010·4
CJS
第30卷

陈云松 范晓光

摘要:因果关系是社会学分析的核心议题。然而,基于调查数据的定量分析都会受到内生性问题的困扰,从而往往难以作出恰当的因果性推断。大量现有的社会学定量分析虽以解释为己任,但实质上要么仅仅停留在描述统计相关阶段,要么得出错误的因果性结论。本文以社会互动研究为例,详细讨论了遗漏偏误、自选择偏误、样本选择偏误和联立性偏误等四种主要的内生性问题的来源,并介绍了克服内生性问题的一系列模型识别方法。最后还基于中国综合调查(CGSS2003)问卷,提出了通过提高调查数据信息量以利克服内生性问题的可能性。

关键词:内生性 因果效应 社会互动

The Endogeneity Problem in Quantitative Analysis: A Review of Estimating Causal Effects of Social Interaction

Chen Yunsong & Fan Xiaoguang

Abstract: Cause-effect relationships are the core area in sociological analysis. However, sociological analysis based on survey data is confronted by the endogeneity problem which plagues causal inferences. Many existing studies aiming at providing explanations for social phenomena either merely describes the statistical associations among variables or arrives at problematic causal conclusions. Focusing on the social interaction studies, this paper addresses the major sources of potential endogeneity biases, namely, the omitted variable bias, self-selection bias, sample-selection bias and the simultaneity bias. Useful model identification

* 作者:陈云松 牛津大学纽菲尔德学院 (Author 1: Chen Yunsong, Nuffield College, Oxford University) E-mail: yunsong.chen@nuffield.ox.ac.uk; 作者 2: 范晓光, 浙江省社会科学院社会学研究所 (Author 2: Fan Xiaoguang, Institute of Sociology, Zhejiang Academy of Social Sciences)

** 作者感谢 Tom Snijders, Peter Hedström, Nan Dirk de Graaf、陈友华、郝大海、李煜等教授与徐建牛、吴桂英、章奇、李志赞和郭茂灿等同对本文的批评和建议。此外,还要感谢《社会》匿名审稿人宝贵的评审意见。文责自负。

strategies for correcting these problems are reviewed. Based on CGSS2003, this paper also discusses how to partially correct for the endogeneity problem through augmenting the volume of survey data.

Keywords: endogeneity, causal effects, social interaction

近几十年来,社会学家们对社会学发展的现状和目的争议重重。其中一个具有代表性的观点就是,尽管定量分析方法不断发展,但大量的社会学实证研究囿于研究设计、数据质量和模型设置,仅仅停留在统计回归的描述阶段,而缺乏基于因果判断的解释能力。以谢宇、温希普(Christopher Winship)、索布尔(Michael Sobel)、摩根(Stephen Morgan)和莫维(Ted Mouw)等为代表的社会学家强调社会学分析必须基于反事实因果关系框架,重视回归分析中的模型设置问题(参见 Morgan & Winship, 2007);另外一方面,以索伦森(Aage Sørensen)、布登(Raymond Boudon)、埃尔斯特(Jon Elster)和赫斯特洛姆(Peter Hedström)为代表的社会学家则侧重强调要把厘清社会机制、社会过程和统计推断结合起来(参见 Hedström & Swedberg, 1998; Hedström, 2005&2008; 陈云松, 2008)。虽然侧重点各不相同,但这些社会学家都强调了解释性机制或者因果推断是社会学分析的目标。此外,从政策研究的角度而言,也只有因果分析才能预测事件将来在什么条件下能发生并得出干预措施中的控制手段,从而为我们制定对策、改良社会提供依据(王天夫, 2006)。

为避免对因果概念进行过多哲学意义上的讨论¹,本文使用的因果关系概念基于反事实框架(counterfactual framework)。这是一个在近年来主流社会学科界共同接受的因果概念。也即,一个影响因子或者干预(treatment)对个体*i*的因果效应,应该是*i*在控制组和干预组中的两个可能的结果状态之间的差异也即 $\delta_i = Y_t - Y_c$ 。然而,正如古希腊的一句名言“人不能两次踏进同一条河流”所说的那样,个体*i*的结果,只可能在一个组中被观察到。这样,对于*i*而言,或者两者都可能是,或者两者之一必然是缺失数据(关于反事实的定义和框架参见 Holland, 1986; Winship & Morgan, 1999; Winship & Sobel, 2004)。谢宇(2006: 44)曾经以大学教育为例来说明该问题。比如,在分析大学教育

1. 较新的关于因果关系的中文文献可参见章奇(2008)对社会科学领域因果概念的全面回顾。

对于个体的收入是否有因果效应时,对一个上大学的学生,我们不可能获得他不上大学的情况的数据。因此,社会学定量分析中只能用平均干预效应(average treatment effect)来替代,也即估算一组大学生(干预组)与一组非大学生(控制组)之间的平均收入差异, $\bar{\delta} = \bar{Y}_t - \bar{Y}_c$ 。但这个替代的前提就是,干预组和控制组必须在其他收入因素上是一致的。也即,两组人的年龄、性别、家庭背景、智商、性格等等的平均值都一样。一旦两组之间在某个变量 E 的均值上不一致(例如性格,往往难以采集数据),且这个遗漏掉的变量本身和解释变量又有关系,那么我们估算出来的结果就是有偏差的,甚至是伪相关的,由此,因果判断就无从谈起。

在回归分析中,人们一般通过控制一系列变量使两个组别具有可比性。但是,总有那些无法被观察的,或者学者未想到的变量导致两组之间不具有可比性。对于回归方程而言,这就意味着解释变量和遗漏误差项出现了相关,不能满足高斯马尔可夫定理(Gauss-Markov theorem),估计参数就会有偏误。在计量经济学领域,这个问题一般称为内生性(endogeneity)问题(Wooldridge, 2002; 2006; Cameron & Trivedi, 2005)。无论是计量经济学还是定量社会学分析,由于绝大多数实证研究都基于非实验性数据,也即无法保证实验组和控制组的相似性,因此所有基于调查数据的实证研究,无一例外都受到内生性问题的困扰。而这些困扰,主要源自一般性的遗漏变量偏误(omitted variable bias)、自选择偏误(self-selection bias)、样本选择偏误(sample-selection bias)和联立性偏误(simultaneity bias)等多个方面。¹既然社会学的研究旨趣在于阐明某种机制或因果关系,而非局限于描述自变量和因变量之间的统计相关,那么内生性偏误就应该成为社会学定量研究中必须直面的问题。在社会学研究领域,尽管对于反事实因果关系、模型识别策略的关注一直存在(Winship & Morgan, 1999; Sobel, 1996; Morgan & Winship, 2007),但是,在大量的实证研究中,内生性问题往往被一笔带过,甚至不加提及。而且,回顾性文献也缺乏对内生性问题的全面梳理,或许唯一的例外可能只有 T. 莫维(Mouw, 2006)对于社会资本因果效应的文献回溯。而中文社会学文献则对内生性问题关注更少。王天夫(2006)和章奇(2008)对社会科学领域因果分析的基本概念进行了非常出

1. 实际上,还有一些导致内生性问题的来源,比如测量误差等等,限于篇幅,本文不加讨论。

色的回顾,但遗憾的是,他们却也没有对与因果推断紧密相关的内生性问题及其解决方案进行深入的探讨,本文则试图弥补这一空白。

在扼要说明内生性问题的来源之后,本文将以社会互动等具体研究为例来讨论解决内生性问题的方法。文中的社会互动(social interaction)是一个非常宽泛的概念,指人们以相互的或交换的方式对别人采取行动,或对别人的行动作出反应,它是人类活动的重要形式(波普诺,[1995]1999:116)。曼斯基(Manski,1993)、杜尔拉夫(Durlauf,2002)和 T. 莫维(Mouw,2006)都指出,社会资本(social capital)、社会资源(social resources)、社会网效应(network effects)、社会规范(social norm)、同群效应(peer influences)、邻里效应(neighbourhood effects)、社会模仿(imitation)、社会濡染(social contagion)等领域的研究实质上都可以归入社会互动这一范畴的名下。尽管,与经济学家相比,社会学家更早认识到社会关系、社会网络、集体行为和特征对个体水平上的结果有重要影响,但经济学家对社会互动领域研究的内生性问题更为重视,在实证研究中大量应用不同的统计模型、精妙的研究设计和丰富的数据以增强模型识别性(model identification),从而推进内生性问题的解决。¹其中被广为运用的方法包括工具变量(instrumental variable)、固定效应模型(fixed effects model)、同胞效应(sibling effects)、倾向评分匹配(propensity score matching)、实验以及准实验(experiments and quasi-experiments)等等。对社会互动文献中解决内生性问题的方法进行梳理,可以为社会学定量分析提供极富价值的参考。

一、内生性问题来源

内生性问题的最主要的来源是模型设置错误和测量误差。本文主要关心的是前者,也即广义的遗漏变量问题。在广义的遗漏变量偏误概念之下,可以把这些内生性问题的来源划分为四个不同的类别。

(一) 一般性遗漏变量偏误

在不同的学科中,遗漏变量偏误经常被称作未观察到偏误(unobservable bias)、隐藏偏误(hidden bias)、紊乱问题(confounding

1. 相关理论文章多由经济学家所完成,参见 Manski, 1993; 2000; Moffitt, 2001; Durlauf, 2002; Durlauf & Fafchamps, 2004; Bramoullé, Djebbari & Fortin, 2007。

problem)抑或未观察到的异质性问题(unobserved heterogeneity)。以社会资本分析为例,一个典型的回归模型可以如下的公式来表示:

$$y = \beta_0 + \beta_1 x_1 + \beta X + \epsilon \quad (1)$$

其中 y 表示劳动力市场结果, x_1 为社会资本变量,大写的 X 为另一个控制项向量, ϵ 则为误差变项。如果 ϵ 与 x_1 不相关,那么我们可以利用普通最小二乘法(OLS)对方程(1)进行无偏估计。然而,如果重要的遗漏变量 x_2 随 x_1 和 y 而变化,那么误差变项必然包含 x_2 ,而且与 x_1 相关。因此,对 β_1 的普通最小二乘估计是有偏的。

那么,这个偏误到底有多大?我们假设“真”模型应该为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta X + \epsilon \quad (2)$$

而且遗漏变量 x_2 与社会资本变量 x_1 的相互关系表示为如下方程

$$x_2 = \delta x_1 + v \quad (3)$$

我们用符号“ \sim ”而不是“ \wedge ”来强调 $\tilde{\beta}_1$ 来自模型(1), $\tilde{\beta}_2$ 则来自模型(2)。因此,我们可以得出 $\tilde{\beta}_1 = \beta_1 + \beta_2 \delta$ 。这样,遗漏变量偏误在估计模型(1)中即为 $\beta_2 \delta$ 。

在社会资本文献中,普遍存在遗漏变量偏误的可能。假设我们要估计“找熟人”(using contacts)相对于“不找熟人”对于求职的因果影响,如果在模型中个体的“口才”被遗漏,那么我们估计出的使用社会关系对工资收入的影响可能就是有偏误的。其原因显而易见:口才不仅增加了找关系的成功概率,而且本身也影响工作类型和工资水平——因为较高的语言技能本身就代表了较强的能力。如果我们分析“使用了的社会资本”(used social capital)的效应,基于同样的道理,那些具有更好口才的求职者就可能找到拥有更高社会地位的帮助者推荐或提供更好的工作。如果分析“可使用的社会资本”(accessed social capital)的个体效应,我们同样可以预期口才好的人可能会交往到更多和更富有社会资源的朋友圈子,同理,被估计的因果效应同样也会存在偏误。

(二) 自选择偏误

作为一种特殊的遗漏变量偏误,自选择问题指的是主解释结果/因变量是在某种程度上被个人选择所决定(Cameron & Trivedi, 2005: 546)。要进一步理解“选择”,我们可以把所分析的社会现象解析为两个过程。比如,一个过程是解释变量发挥作用的主体过程,而另外一个

则是个人选择的过程。与一般性的遗漏变量偏误相比,自选择问题的实质就是两个过程的非观测因素相互关联。

现在我们针对两个过程分别列出两个模型:

$$y = \beta_0 + \beta_1 x_1 + \beta X + \epsilon \quad (4)$$

该模型与方程(1)完全相同。而对于选择模型,我们将个体进行选择效用(U)定义为解释变量 x_1 的函数,观察的个体特征 X' (X' 与 X 以示区别)和个体本身所知而不为社会学家所证实的 ϵ 的向量。因此我们可以得到如下方程:

$$U = F(x_1, X', \epsilon') + \eta \quad (5)$$

假设同时影响两个过程的因子 x_2 是未被社会学家所发现的非观测因素,因此在方程中被遗漏。换言之, x_2 进入 ϵ 和 ϵ' 。既然方程(5)规定了 ϵ' 和 x_1 相关,那么 x_2 和 x_1 之间必然存在关联。因此, x_1 和 ϵ 具有关联性。这样,基于方程(4)所得出的估计 β_1 就会是有偏误的。大量研究已经证明,选择性行为是社会资本过程中非常重要的因素。以下笔者用“找熟人”和“朋友们的平均地位”两种典型的社会资本变量来说明这个问题。

在社会资本研究领域,学者可能对“找熟人”相对于“不找熟人”的求职效应更为关注。但研究发现,拥有更高地位、更长受教育年限和更丰富社会资本个体可能会倾向于较少地使用社会关系(Marsden, 2001; Bentolia *et al.*, 2003; Lin & Ao, 2008)。如果一个求职者因为自身的某个特点选择了“找熟人”这种求职方式,那么“找熟人”事件本身对于他求职的影响,就很难说是因果性的。在这里,我们把决定是否“找熟人”(x_1 为虚拟变量)定义为被求职者自己觉察到的效用方程。比如,假设“胆怯”是 x_2 ,也即一个同时影响找关系的决策(x_1)和工资水平(y)的遗漏变量,那么“胆怯”势必进入 ϵ 和 ϵ' 。既然方程(5)设定“胆怯”与“找熟人”(x_1)相关,那么方程(4)里的偏误项 ϵ 就自然也会与 x_1 相关了。因此,方程(4)产生了偏误估计。

与此相似,根据社会趋同(social homophily)理论,社会网在种族、性别、社会阶层、宗教信仰、行为和价值观等方面具有更大的选择性,也即“物以类聚,人以群分”(McPherson *et al.*, 2001)¹。若是人们自己选

1. 在社会学文献中,对自选择的关注由来已久。参见 Duncan *et al.*, 1968, Blalock, 1984, Kandel, 1978.

择了具有某种特点的朋友,那么被观察到的朋友特征对于其个体结果的影响也未必是因果性的了。这是因为社会资本与个体结果间的相关性也许仅仅是反映了组群内大家所共同具备的一些选择性因素(参见 Mouw, 2003; 2006)。这种自选择在经济学的社会资本研究中又被称为“内生性组群构成(endogenous group formation)”(Moffit, 2001)。

(三) 样本选择偏误

当因变量的观察仅仅局限于某个有限的非随机样本时,样本的选择偏误就容易产生。这种对某些观察值的非随机性排斥(exclusion)不仅源自数据收集程序,而且也来自于研究中社会现象本身所固有的特质。在经济学领域,赫克曼(Heckman, 1976)仔细探讨了非随机样本造成的选择偏误问题。他事实上将其作为一种特殊的遗漏变量偏误。Berk (1983)最先在社会学文献中介绍了样本选择偏误。

假设社会学家想估计对于 y 在总体样本上的影响,不过有时 x_1 仅在 $S = 1$ 的情况下才能被观察到(例如, x_1 表示帮助求职者的资源,而 x_1 仅在使用了社会关系的情况下才能被观察),则方程往往被写成:

$$y = \beta_0 + \beta_1 x_1 + \beta X + \epsilon \quad (6)$$

但实际上对于总体样本我们只是估计了

$$Sy = S\beta_0 + \beta_1 Sx_1 + \beta SX + S\epsilon \quad (7)$$

要从方程(7)获得 β_1 的无偏估计,就必须满足 $S\epsilon$ 与 Sx_1 不相关。既然 S 是二分变量,事实上,这个条件就是 ϵ 必须与 Sx_1 不相关。

现在我们将选择因子 S 定义为:

$$S = F(X', \epsilon') + \eta \quad (8)$$

在方程中, X' 是可被观察到的控制变量,而 ϵ' 则为社会学家无法观察的选择因素。 F 是决定选择指数 S 等于 1 或 0 的函数。显然, X' 和 X 有可能部分重叠,而且 ϵ' 和 ϵ 也一样。一旦有某个遗漏因子 x_2 同时进入 ϵ' 和 ϵ , 方程(6)的估计就一定是偏的。方程(8)实际上与方程(5)很类似,这也正是样本选择偏误和自选择偏误都可以被叫做选择偏误(selection bias)的缘由。

其典型的例子是估计“使用了的社会资本”(used social capital)的因果效应。要调查该社会资本的效应,最直观的方法是对表示自己的工作“通过协助者找来的”被访者进行分析,观察他所“动员”或“使用”到的资源量(如求职协助者的职业声望)是否对收入有显著效应。

然而,某些未被观察的变量(如自尊心)可能同时作用于个体工作结果和使用社会关系求职的选择性之上——自尊心强的人不愿意找人求职,而且自尊心导致工作努力和较高的收入。若仅对那些报告了使用社会关系的局部样本建立简单的 OLS 模型,这样等于只对一个自尊心不强、工作不努力的样本进行估计,社会资本的因果效应可能由此被高估。

(四) 联立性偏误

联立性问题(simultaneity)也是内生性的一个重要来源。其本质就是解释变量连带地由因变量决定,也即双向因果关系。例如,在“可使用的社会资本”(accessed social capital)研究中,我们总是把个体当前职业声望得分对被调查者社会网络的平均职业声望进行回归。然而,由于社会网的声望均值显然同时也受到被访者个体声望的影响,用标准的均值线性模型就会产生很多问题。本文结合曼斯基(Manski, 1993)和莫菲特(Moffit, 2001)的研究,作一个简要的阐释。

在不失去一般性的情况下,估计“可使用的社会资本”效应的均值线性(linear-in-means model)模型可以表示为:

$$y_{ig} = \beta_0 + \beta_1 \bar{y}_{-ig} + \beta_2 \bar{X}_{-ig} + \beta_3 X_{ig} + \epsilon_{ig} \quad (9)$$

在方程(9)中, i 代表个体, g 代表组群, x 代表个体特征(如性别、教育获得等)。需要注意的是, \bar{y} 为组群结果的均值,下标 $-i$ 表示排除个体自身计算所得的均值, \bar{X} 为平均组群特征。其实, β_1 和 β_2 都可被视为社会资本效应(Durlauf, 2002)。

为了对此作简明表述,假设每个组群只有 2 个个体($i=1, 2$),模型(9)可以写为:

$$y_{1g} = \beta_0 + \beta_1 y_{2g} + \beta_2 X_{2g} + \beta_3 X_{1g} + \epsilon_{1g} \quad (10)$$

$$y_{2g} = \beta_0 + \beta_1 y_{1g} + \beta_2 X_{1g} + \beta_3 X_{2g} + \epsilon_{2g} \quad (11)$$

如果将方程(10)右边插入方程(11),我们可以得到简化形式

$$y_{2g} = \frac{\beta_0}{-\beta_1} + \left(\frac{\beta_1 \beta_2 + \beta_3}{1 - \beta_1^2} \right) X_{2g} + \left(\frac{\beta_1 \beta_3 + \beta_2}{1 - \beta_1^2} \right) X_{1g} + \left(\frac{\beta_1 \epsilon_{1g} + \epsilon_{2g}}{1 - \beta_1^2} \right) \quad (12)$$

然而,这也带来了一些问题。第一, y_{2g} 必然与 ϵ_{1g} 相关,因为 y_{2g} 与 y_{1g} 相关。结果,来自方程(12)的估计是有偏的。这就是联立性偏误问题。第二,即使 ϵ_{1g} 与 y_{2g} 不相关,只有 $\left(\frac{\beta_1 \beta_2 + \beta_3}{1 - \beta_1^2} \right)$ 和 $\left(\frac{\beta_1 \beta_3 + \beta_2}{1 - \beta_1^2} \right)$ 可以被识别,而且无法将 β_1 和 β_2 分离。后者被曼斯基(Manski, 1993)称为“映

射问题”(reflection problem)¹。

二、模型识别的策略选择

内生性问题和模型识别(model identification)紧密相连。在微观计量经济学的术语中,模型就是对一系列观察值概率分布的设置(specification),而结构(structure)则是对该分布参数的设置。假如两个及以上模型指向相同的联合概率分布,那么这两个结构是等相同的,也就是无法识别的。当模型估计的目的是为评估解释变量的因果系数时,得出无偏估计的重要的前提条件是同构的替代模型应被剔除(Cameron & Trivedi, 2005: 20 - 30)。一般而言,可以把识别某个模型理解为排除竞争性解释存在的可能,也即控制那些无法观测或者被遗漏的且和解释变量相关的因子,故解决内生性问题的术语就是模型识别。本节将针对四种内生性问题来源分别讨论相应的模型识别策略。此外,模型识别必须以包含足够信息量的数据为基础,因而本节也会专门就提高调查数据信息量做一简要论述。

(一) 解决遗漏变量偏误

在现有的实证研究中,主要有四种策略用于消解遗漏变量偏差问题。

第一是利用非传统数据作为遗漏变量的替代(proxy)以控制潜在的遗漏变量。例如,在邻里研究中,詹克斯和 S. E. 迈尔(Jencks & Mayer, 1990)证实,当纳入更多对父母特征的控制时,生长在贫困社区这一因素对于个体的影响趋于下降。相类似的,G. J. 邓肯、布鲁克斯-冈恩和克列巴诺夫(Duncan, Brooks-Gunn & Klebanov, 1994)在模型中纳入了对家庭温暖和家庭学习氛围的控制变量。在学校效应研究中,克拉克和诺埃亚克(Clark & Youenn Lohéac, 2005)控制了父母是否选择社区。然而,我们永远不知道是否所有未被观察的因子都被控制。既然建立起“穷举式”(exhaustive method)的模型在理论上是不可能的,一个替代性方案则是利用滞后因变量(lagged dependent variable)作为未被观察的个体异质性和历史因子的代表。例如,伍德里奇(Wooldrige, 2006: 315)指出,这种方法对横截面研究(cross-sectional study)的数据

1. 如此命名的原因主要是这两者就像人和镜子里映射的影子一样同时变化和出现,让外面的观察者无法分析谁是原因,谁是结果,谁是主因,谁是次因。

条件提出了更高的要求,而且它也提供了“一种处理导致因变量现有差异的历史因子而且又迥异于其它处理方式的简单方法”。

第二种方法依赖于变量在时间维度上的差异,运用固定效应模型去控制“时间固定的”(time-invariant)的非观测因素。其在社会学中的应用可参见哈拉比(Halaby, 2004)的综述。在学校的同群效应研究中,哈努谢克等(Hanushek *et al.*, 2003)利用 Texas 学校的面板数据(panel data)控制了固定的个体、学校年级以及学校效应,由此发现了学校年级同群(school by grade)的平均成绩对个人成绩的增长有正向作用。在社会濡染研究中,伊基诺和 G. 马吉(Ichino & Maggi, 2000)控制了个体和银行分支机构的固定效应,以此来研究银行分支机构雇员的罢工行为是如何被分支机构的平均罢工行为所决定的。不过,正如阿伦森(Aaronson, 1998)所指出的,固定效应方法往往会得到较大的标准误差。更重要的是,它无法控制随时间、家庭、学校变化的非观测因素。

在实证研究中,对同一被访者进行重复观察的“真”面板数据往往难以收集。不过,在特定假设下,组织的管理记录可以用来追踪一段时间内的变量差异,当然也就可以进行固定效应分析(fixed effects analysis)。例如,利用 Texas 学校微观数据样本,霍克比(Hoxby, 2000)用校内班级年度变化(within-school year-to-year class variation, 也即学校内同一班级的跨同期群或跨年度的比较)来估计班级成员的构成对于学生考试成绩上的影响。因为校内班级的年度变化可以被假设为随机的,所以估计参数可以被视为是无偏的。类似地,古尔德等(Gould *et al.*, 2005)利用同一年级里移民数量的跨年度变化来估计以色列学校中移民比例对学生高中考试成绩的影响。V. 拉维等(Lavy *et al.*, 2008)运用留级生-跳级生-正常生(repeaters-skippers-regulars)在班级中比例的随机变化来估计班级结构对学生学业成绩的影响。

第三种应对遗漏变量偏误的方法,是利用组内策略(with-group strategy),也即用非时间性(non-time)的组内差异去估计固定效应模型。这尤其适用于处理遗漏变量存在于组别水平(group-level heterogeneity)上的情况。如在许多研究中,组别水平的异质性是主要的内生性来源,因为父母可能会择校,某些无法观察的学校特征会成为影响估计班级效应的干扰项。然而,如果同一学校的班级可以被假定是随机分配的,那么比较校内各班级的差异就可以把学校效应从模型

中排除。阿莫米勒和皮施克(Ammermüller & Pischke, 2006)利用校内差异估计了欧盟国家六个小学的班级对于学生个体考试成绩的同群效应。麦克尤恩(McEwan, 2003)用同样的方法分析了智利中学的年级效应。而阿西迪亚科诺、尼科尔松(Arcidiaconor & Nicholson, 2005)则阐释了医学院中的班级平均能力和班级专业选择是否影响学生的个体学业成绩以及专业选择的机会。

除了学校效应、班级效应的研究,这种组内策略同样被运用到邻里效应、社会网络效益等研究中,以解决潜在的遗漏家庭特征变量问题。该方法的根本逻辑是,通过比较在不同社区或朋友圈子的同胞兄妹,以此消除家庭固定效应,从而准确地评估邻里或者朋友圈的同群效应,故它也被称为“家庭固定效应”(family fixed effects)或“同胞固定效应”(sibling fixed effects)方法。利用此方法,阿伦森(Aaronson, 1998)发现邻里教育获得与个体完成高中学业的倾向之间存有显著正向关系;帕洛尼等(Palloni *et al.*, 2001)证明社会资本影响墨西哥人的移民决策;其它采取类似方法的社会互动研究则见于青少年社会网、风险行为等领域(Soetevent & Kooreman 2005, Clark & Youenn Lohéac, 2005)。

第四种处理遗漏变量偏误的方法是使用工具变量(若能发现并使读者相信某个外生因素与误差项无关但又和解释变量高度相关,则其或许是一个好的工具变量)。如前所述,家庭特征在识别邻里、班级或学校同群效应时往往是遗漏变量。为了解决该问题,社会学家常常把州、县或大都会地区层面的汇总数据作为解释变量的工具变量(Card & Krueger, 1996)。例如,埃文斯、W. E. 奥茨和施瓦布(Evans, Oates & Schwab, 1992)运用大都会地区的失业率、家庭收入中位数和贫困率作为学校中贫困学生比例的工具变量,他们发现,学校中贫困生比例对学生怀孕或辍学行为并无显著的同群效应。卡尔特和克莱泽(Culter & Glaeser, 1997)则把地方政府辖区分隔及贯穿大都市的河流数量作为邻里区隔(segregation)的工具变量。他们发现生活在黑人高集中区域的黑人比其他地方的黑人更具劣势的社会经济结果。西波隆和罗索利亚(Cipollone & Rosolia, 2007)以地震导致的男性免于征兵的政策来作为高中班级性别构成的工具变量,以分析意大利学生中班级性别构成对女生的影响。其他利用工具变量方法来估计同群效应的研究可以参见G. 邓肯、康奈尔和克列巴诺夫(Duncan, Connell & Klebanov, 1997)及

居里和叶洛维茨 (Currie & Yelowitz, 2000) 等人的研究中。

在劳动力市场研究方面, 邦托利阿、米凯拉奇和苏亚雷斯 (Bentolia, Michelacci & Suarez, 2003) 使用两个工具变量如年长的兄姊数目和联邦就业率来解决“使用社会关系”进行求职的内生性问题, 发现因工人职业与能力的不匹配, 使用社会关系反而导致低工资。孔特雷拉斯等 (Contreras *et al.*, 2007) 借鉴巴耶尔和 S. 罗斯 (Bayer & Ross, 2006) 的方法, 用房价来控制邻里未被观察因素, 用和被访者具有同样可观察特征者的平均邻里特征来作为邻里特征的工具变量, 发现邻里的非农就业对玻利维亚妇女获得非农工作具有正面影响。E. 莫林和莫斯基翁 (Maurin & Moschion, 2009) 则考察了法国邻里的劳动力市场参与是如何影响母亲入职的决定。为了消解内生性问题, 他们运用邻里平均的儿童性别混合度 (sex mix) 去作为邻里劳动力市场参与的工具变量, 结果证明邻里就业和母亲个体就业息息相关。然而, 如 P. J. 哈丁 (Harding, 2003) 评述的那样, 工具变量评估往往有更大的标准差。更为重要的是, 工具变量是否与误差项无关联并不可验 (untestable)。特别是, 汇总数据作为工具变量反而往往会增加遗漏偏误 (Grogger, 1996, Hanushek, Rivkin & Taylor, 1996, Moffit, 1995, Rivkin, 2001)。

(二) 解决自选择偏误

在实证研究中解决自选择偏误, 最直观的方法就是设法使得选择行为不存在。因此, 随机分配 (Random Assigning to Treatment) 是解决问题的最佳途径。这是因为, 随机分配可以确保主解释变量和未观察因子之间没有任何关联性的出现。实验和自然实验 (natural experiment) 可以设计出随机分配, 它们显然是理想的方法。

在干预效应 (Treatment Effects) 研究中, 迪弗洛和 E. 赛斯 (Duflo & Saez, 2003) 运用实验来探究群体信息传播对个人退休计划决定的影响。雅各布 (Jacob, 2004) 分析了芝加哥地区公共住房拆迁对拆迁户学生标准化考试成绩的影响。霍克比、魏因加特 (Hoxby & Weingarth, 2005) 用维克 (wake) 县学区的再分配政策作为自然实验。基于面板数据的分析, 他们认为该县的学区再分配计划确实影响了学生的成绩。最著名的实验性社会互动研究是基于美国“联邦流动机会” (MTO) 计划的一系列文献。通过此类随机实验, G. 邓肯等 (Duncan *et al.*, 2004) 发现迁入富裕社区对于成人的社区邻里安全感和精神健康有相当大的改

善。桑布曼特苏等(Sanbonmatsu *et al.*, 2006)发现了 MTO 计划对迁入新社区的孩子的阅读水平和数学考试成绩没有显著性影响。卢德维格、G. 邓肯和平克斯通(Ludwig, Duncan & Pinkston, 2005)则指出,仅仅给家庭提供居住迁移机会对家庭的社会经济结果影响微乎其微。克林格、卢德维格和 L. F. 卡茨(Kling, Ludwig & Katz, 2004)发现,迁入贫困社区反而减少了女青年的暴力行为和财产犯罪。除了这种“田野”实验,还有一些研究运用了“室内实验”,具体可参见福尔克和菲斯巴赫尔(Falk & Fischbacher, 2002)及 T. N. 卡森和 V-L. 缪(Cason & Mui, 1998)等的研究。

除了干预效应研究,实验或自然实验也同样适用于研究群体对个体的影响。例如,萨塞尔多特(Sacerdote, 2001)利用达特茅斯(Dartmouth)大学大一新生宿舍随机分配这个自然实验,发现了作用于学术成就、GPA 等方面的同群效应。齐默尔(Zimmerman, 2003)利用威廉姆斯学院的一年级宿舍随机分配制度,分析了室友原先 SAT 成绩对于个体成绩的影响。另外,在社会互动研究中,外生性干扰(如自然实验,政策介入或自然发生事件)可以用来建构排除限定(也即工具变量)以帮助识别模型。如利伯曼、L. F. 卡茨和克林格(Lieberman, Katz & Kling, 2004)考察了人口普查中个人结果(如风险行为、精神健康、基于社会性别的生理健康)的邻里效应。不过,随机分配的实验和自然实验也有缺点,即它们往往是费钱费时甚至不可行。另外一个值得提及的是,诸如 MTO 的社会实验很可能其本身就包含样本选择偏误。如 MTO 项目只针对那些处于一定贫困线以下的人,结果绝大部分参与者是黑人和拉丁裔人口(Sampson, 2008, Clampet-Lundquist & Massey, 2008)。

第二种校正方法是匹配(matching),其被广泛应用于干预效应研究。不过,该方法的前提假设是,自选择完全由观察到的变量所导致。罗森鲍姆和 D. B. 鲁宾(Rosenbaum & Rubin, 1983)是该方法的首创者,并将其命名为倾向评分方法(PSM)¹。其直观的理念是,用尽量相似的控制组成员和对照组成员来进行比较,以减少偏误。例如, S. L. 摩根(Morgan, 2001)研究了选择天主教学校对于学习成绩的干预效应,

1. 倾向评分是指基于个体的可观察特征而进行某种选择的条件概率。运用 PSM 的主要步骤有: 1) 用 logit 和 probit 方程对选择建模,并且把估计到的概率作为个人倾向得分; 2) 为每个选择了的被访者匹配一个非常接近的不选择、或不参与的人。接近标准就是倾向得分; 3) 比较匹配对子的结果变量。

他运用 PSM 估计去弥补回归模型所得到的估计。为了探究邻里选择对高中辍学和未成年人怀孕的效应, D. J. 哈丁(Harding, 2003) 引进 PSM 方法来比较处于青春期的具有相同观察因素和不同邻里的两组儿童。他发现那些较高贫困型社区的个体比较低贫困型社区的高中生更有可能发生辍学和未成年怀孕。然而, PSM 的两个缺陷是它非常依赖于大样本和对照组、控制组匹配重叠程度(Morgan & Harding, 2006)。而且, 它对未被观察的干扰项无能为力¹。不过, PSM 可以和工具变量等方法结合起来使用进行相互比对, 有兴趣的读者可参见迪普雷特和甘格尔(DiPrete & Gangl, 2004)。

第三种方法是建立选择过程的模型。建立基于选择模型和实质模型的联立方程组是一种校正自选择偏误的简明方法。如安格里斯特和拉维(Angrist & Lavy, 1999)运用回归间断设计(regression discontinuity design)建立在实验中由班级规模引起的班级拆分可能性模型。尔后, 他们检定了班级规模和学生考试成绩之间可能的因果关系。此外, 其他的研究则运用离散选择模型和其他计量经济学方法来分析邻里选择、教育消费和决策(Bayer, McMillan & Rueben, 2002; Conley & Udry, 2008; van der Klaauw & van Ours, 2003)。赫克曼(Heckman)选择估计方法同样也是一种对参与决策过程进行建模的处理自选择偏误方法。如德拉特、萨巴蒂耶(Delattre & Sabatie, 2007: 223)就曾比较了三种模型(OLS, 赫克曼二阶段[two-stage]法, 以及内生性切换回归([endogenous switching regression]))的结果, 以探究“使用社会关系”对工资的直接作用。不过, 赫克曼法主要适用于下文即将提到的“样本选择偏误”。该方法在解决自选择研究中运用有限, 这主要是由于可供选择的其它方法很多, 诸如实验、工具变量估计和面板模型等(Bushway *et al.*, 2007)。

最后, 由于自选择问题也可以被看作是一种遗漏变量偏误的特殊类型, 因此我们也可以通过纳入更多的控制变量、代理变量和固定效应模型等传统方法来处理这个问题。哈特曼和伊尔迪茨(Hartmann & Yildiz, 2006)通过把众多体现个体异质性的变量纳入模型, 以此解决内生性组群构成问题。克劳特(Krauth, 2004)用个体对观察到的变量选择程度

1. D. J. 哈丁(Harding, 2003)在进行敏感性分析(sensitivity analysis)之后为 PSM 方法进行了辩护, 认为“无法观察因子对社区邻里和结果之间的关联性难有较强贡献”。

来作为未被观察变量的自选择的替代变量,以此研究加拿大青年中同群在吸烟行为上的效应。为了识别班级同群对学生成绩的效应,卡曼和张磊(Carman & Lei Zhang, 2008)指出,在中国,班级选择和教师分配在中学入学时就确定,而且此后几乎不变。因此,他们将固定效应模型应用到面板数据中,将选择行为作为时间固定的遗漏变量加以排除。

(三) 解决样本选择偏误

被广泛运用来解决样本选择问题的方法是赫克曼(Heckman, 1976)的二阶段法。赫克曼用一个类似于前文中方程(8)的函数来代表选择过程:

$$s = 1 \text{ if } \gamma_0 + \gamma_1 x'_1 + \gamma_2 x'_2 + \cdots \gamma_m x'_m + \epsilon' > 0 \quad (13)$$

第一步,利用总体样本里所有的观察值,以 s 为被解释变量, x 为解释变量估计 probit 模型,由此获得一系列参数估计值 $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_3)$ 。对每个观察值 i 计算反向 Mills 比率(或风险比率) $\hat{\lambda}_i$ 。R. A. 伯克(Berk, 1983)指出,线性概率模型的风险比率等于 $s=0$ 的估计概率减去 1.0,而来自 logit 模型的风险比率就是 $s=0$ 的估计概率。第二步,再利用选择样本($s=1$)进行回归,此时的自变量应该包括 $(x_1, x_2, \cdots, x_k, \hat{\lambda})$,也即把代表样本选择的风险比率纳入方程 9。这样便可以获得一致估计量。赫克曼二阶段法在社会科学界得到广泛应用,主要是因其简明性及 OLS 回归技术在研究中的普遍运用(Hughes, 1997)。当然,该方法也有若干问题。施托尔岑贝格和雷利斯(Stolzenberg & Relies, 1997)提出了判定赫克曼二阶段估计究竟是增加还是减少了回归系数估计准确性的方法。布沙威等(Bushway *et al.*, 2007)强调了赫克曼法应用的四种常见问题,包括二分因变量的误用¹,计算反向 Mills 比率的难度,错估标准误差,以及真实模型的修正项与其他回归量之间的共线性。

在社会资本的劳动力市场研究中,P. 马登和赫尔伯特(Marsden & Hurlbert, 1988)最早使用赫克曼二阶段法来解决样本选择问题。他根据 R. A. 伯克(Berk, 1983)及 R. A. 伯克和 S. C. 瑞尔(Berk & Ray, 1982)的方法,运用 logit 模型去预测个体使用社会关系的可能性,尔后

1. 例如,如果我们想要检验,使用社会关系是否对工作满意度或者其他形式的离散变量产生影响,那么采用 Heckman 的二阶段估计法来减少选择偏误的做法是有问题的。

将使用社会关系可能性的预测值作为一个控制项纳入到主体工作模型。之后,韦格纳(Wegener, 1991)也使用了同样的方法。较近的研究如林南(Lin, 2004)也运用 probit 和 logit 模型去预测使用社会关系的倾向(propensity),尔后控制其预测值以消解样本选择偏误。不过,正如伍德里奇(Wooldridge, 2006: 620)所强调的,要真正解决样本选择问题,主体模型中的解释因子群应该是选择模型的解释因子群的完全子集,质言之,出现于主体模型中的全部自变量应该都被纳入选择方程。更重要的是,应该在选择模型中至少有一个特殊的解释变量,而它不能出现在主体模型之中。对于社会资本研究而言,也就是要有一个变量只对个人使用社会关系的倾向有影响,而不影响工作结果。没有这个特殊变量,赫克曼法实际上就很难解决样本选择问题。遗憾的是,在上述林南的研究中,出现在选择模型而非工作结果模型中的变量是与教育相联的工作经验和父亲工作的单位级别。上述 P. 马登和赫尔伯特研究中的这个特殊变量则是工作经验。然而,人们有充分的理由去怀疑,这些变量(如工作经验)应该同样也出现在主体模型之中¹。也即,排除限定并非完美,因此在这些应用中还存在估计偏误。

由于样本选择也可以被视为遗漏变量问题的特殊类型,工具变量方法实际上也可以对它加以处理。例如,克林格、利布曼、F. 卡茨和杉波玛茨乌(Kling, Liebman, Katz & Sanbonmatsu, 2004),克林格、利布曼和 L. F. 卡茨(Kling, Liebman & Katz, 2007),以及克林格、利布曼和 L. F. 卡茨(Kling, Liebman & Katz, 2004)考察了 MTO 项目中那些利用 MTO 迁移项目对于迁移人员的干预效应,也即被干预样本中的干预(TOT)效应(如犯罪率,成人经济上的自我满足和身体健康)。在这方面,样本选择显然是个问题。这是因为,尽管提供迁移凭证是随机的,但是那些获得了迁移许可的人员中总有一部分并没有迁移。因此,调查样本一定是那些有资格迁移人口的一部分。如果迁移决策是内生的,那么估计 TOT 必然有偏。所以,他们利用是否获得 MTO 迁移许可作为最终迁移决定(也即主解释变量:是否使用了迁移许可)的工具变量。这个工具变量之所以有效,是因为人们是否能得到 MTO 凭证是随机分配的,故它与影响个体结果的未被观察的个体因子无关,而显然与是否使用凭证的决策正向相关。

1. 例如,劳动经济学中著名的 Mincer 工资方程就把经验作为预测工资的主要解释变量之一。

(四) 解决联立性偏误和映射问题

我们首先关注解决映射问题的方法,然后再转入解决联立性偏误的策略。曼斯基(Manski,1993)在一篇经典论文中指出,有两种不同类型的社会互动效应:(1)内生互动效应(endogenous effects,也即方程[9]中的 β_1):即组群结果(\bar{y})影响个体结果 y ;(2)外生互动或情境效应(exogenous or contextual effects,也即方程[9]中的 β_2):即组群特征(\bar{x})影响个体结果 y 。以劳动力市场为例,假若求职者具有一个社会网络,在其他条件不变的情况下,当他的收入受到朋友平均收入的影响时,被称之为内生互动效应;当他的收入受到朋友平均教育水平的影响时,则被称之为外生效应。恰如前文所述,由于映射问题的存在,人们常常难以区分外生性效应和内生性效应。

如果映射问题对于研究目的是无关紧要的话,那么大可置之不理(Araujo *et al.*, 2004, Hanushek *et al.*, 2003, Arcidiacono & Nicholson, 2005)。例如,阿莫米勒和皮施克(Ammermüller & Pischke, 2006)在分析班级对于学生个体的同群效应时,就明确说明没有区分这两种效应,而是报告了一个“复合”估计量。如果研究目的是区分内生和外生社会互动,映射问题则可用一些简单的方法加以规避。例如,曼斯基(Manski, 2000)认为运用中位数而非平均数来作为组群解释变量可以解决这个问题。另一种方法则是曼斯基(Manski, 2000)所说的利用主观数据(subjective data)。例如,运用1997年有关青年的国家纵贯调查数据,川口大字(Kawaguchi, 2004)把青少年对同群行为的主观认知作为解释美国青年吸毒的解释变量。此外,把非线性引入模型,也将有助于识别两种效应(Manski, 2000)。布洛克和杜尔拉夫(Brock & Durlauf, 2001)已经证明映射问题可以在非线性模型中加以避免。这说明,在那些有关作用于诸如劳动力市场参与及工作满意度等非连续因变量的社会资本效应研究中,因为我们使用logit或者probit模型,映射问题就可以自然避免。近来,布拉穆莱等(Bramoullé *et al.*, 2007)证实,如果个体的组群规模不同(比如,每个人朋友数量不同),那么哪怕是线性均值模型,内生和外生互动效应大体上也可以被识别。

关于联立性偏误的纠正方法。首先检视一些特殊情况:即人们可以假定由相互因果关系引致的联立性是可以不予考虑的。例如,H. 奈尔等(Nair *et al.*, 2006)假定内科医生受到他们心目中的意见领袖(有较

高威望的内科医生)影响,但意见领袖却不受一般的内科医生行动的影响。同样,索伦森(Sorensen, 2005)研究了在普通雇员健康计划选择上的社会互动。他假设新雇员受到老雇员的影响,而并非反之亦然。在L. F. 卡茨(Katz *et al.*, 2001)的研究中,他们假定社区里的老住户不受新迁入家庭的影响。其原因是,迁入家庭数量和原有家庭数量之间的比率很小,也就是说,社会互动的因果效应被假设为单向度的(unidirectional)。既然是单向因果,就无需考虑联立性偏误。

另外,人们还可以强加一个时间序列去打破自变量与因变量之间的联立性。曼斯基(Manski, 1993)实际上把它定义为“利用方程的递归结构(recursive structure)”。也就是说,人们可以假定个体在 $(t+1)$ 时的行为实际仅仅受到 (t) 时组群行为的影响。在实验室的实验中,它比较容易实现。然而,在基于观察数据的实证研究中(Betts & Zau, 2004, Hanushek *et al.*, 2003, Vigdor & Nechyba, 2004),情况要更为复杂。比如,哈特曼等(Hartmann *et al.*, 2008)和苏特温特(Soetevent, 2006)告诫到,使用该方法必须假设未被观察的变量不随时间变化,而且行动者们(agents)必须不具前瞻性。哈努谢克等(Hanushek *et al.*, 2003)用滞后同群变量作为解释变量(explanatory variable)现值的一个代替(proxy)。但是他们也不得不承认,如果误差项连续相关,估计量就会下偏。该方法的另一问题是,根据曼斯基(Manski, 1993)的观点,我们必须假设社会互动效应的实现是按照某个假定的时间序列模式,但是几乎没有任何实证研究说明这个时间有多长¹。

总之,联立性偏误实际亦可看作是一种特殊的遗漏变量偏误。只不过它是由双向因果所导致的。因此,工具变量方法必然是适用的。换言之,至少要找到一个外生变量,它不出现在社会学家进行回归的互动效应方程中,但出现在代表反向因果的模型中。近来有研究常采用外生性政策的干预或者对部分样本的外生干扰来充当工具变量。该方法在本质上与苏特温特(Soetevent, 2006)所称的“偏总体试验”(partial-population experiments)或莫菲特(Moffit, 2001)所提出的“偏好选择干

1. 更确切地说,我们没有关于社会资本效应将持续多久的先验知识。即使理论上可以对两个时间序列进行格兰杰(Granger)因果检验以发现相关变量的后滞期(lag length),但在社会网络研究中,收集这些数据是非常难的。

预”(preference changing intervention)是一回事。如 M. 布泽和卡乔拉(Boozer & Cacciola, 2001)把班级中先前曾经参与过“小班实验”的同学比例作为班级平均成绩的工具,以分析班级平均成绩对个体学业成绩的同群效应。该工具变量之所以很理想,是因为它源自随机设计实验:以前学校从各个班级随机抽人去组成小班,而小班教学提高了这部分学生的成绩。因此,这个数量比与学生个体或家庭的异质性没有关系,却影响平均班级学业成绩。类似方法还可参见博伯尼斯和 F. 菲南(Bobonis & Finan, 2008)对墨西哥教育选择的研究。他们把村庄中参与 Progresá 项目(国家给生活困难的母亲以补助)的比例作为村庄儿童平均就学的工具变量,以此来分析乡村同龄人的入学率是否影响个体的入学决定。一些研究甚至利用环境污染等自然灾害作为外生的随机干扰来构成工具变量。阿萨德尔阿和乔杜里(Asadullah & Chaudhury, 2008)利用村庄中家庭水井里测出砷污染的比例为工具变量,分析了孟加拉国农村八年级学生班级数学考试对于个体数学成绩的内生社会互动效应。其理由是,砷污染会影响儿童智力发育,从而外生地影响那些被污染水井家庭的孩子的学习成绩。

更新的研究提出了一种所谓“条件方差限定”(conditional variance restrictions)的模型识别策略。该方法由格莱塞尔、萨塞尔多特和沙因曼(Glaeser, Sacerdote & Scheinkman, 1996)开创,但由格拉哈姆(Graham, 2008)将其正式模型化。其基本思想如下:对一个既定的社会互动现象,个人结果的跨组别方差(variance)由三个因素所导致:组群异质性的方差、个体异质性的方差以及组群解释变量的方差。如果每个组群由大小不同的次级组群随机组成,那么显然对这些次级组群而言,上级组群和个体的异质性方差是一样的,但它们各自的解释变量的方差却不一样。那么,通过比较不同次级组群的解释变量方差,就可以识别解释变量的因果效应。例如,为了检验幼儿园学业成绩的学校同群效应(school peer effects),格拉哈姆(Graham, 2008)利用 Tennessees 小班实验(STAR 项目),将老师和学生随机分配到两种类型的教室:小班和大班。由于每个小班平均成绩的方差要大于大班,那么所有小班组成的次组群的平均成绩方差要大于大班组成的次组群。这样,班级大小就提供了识别外生性变化的资源。本质上,条件方差限定方法相当于寻找只通过影响组内方差才影响跨组群方差的工具变量(Galibiaty &

Zanella, 2008)。杜尔拉夫和塔纳卡(Durlauf & Tanaka, 2008)称之为对偏误的协方差限定(covariance restrictions on errors),以和一般回归分析中用排除约束加以区别。

(五) 数据收集问题

总体来看,能够部分解决内生性的方法主要是固定效应方程和倾向性评分匹配,而理论上可以全部解决内生性的方法则包括工具变量以及自然实验。更重要的是,固定效应方程、匹配以及工具变量分析,都离不开高质量的调查数据。因此,没有高质量的数据,即使掌握了相对复杂的模型和识别策略,统计分析者也没有用武之地。人们要在定量资料收集和测量上为解决内生性问题尽量提供更多的分析信息,因此,数据质量实际上是解决社会学定量分析内生性问题的重要前提。

例如,2003年中国综合调查(CGSS2003)中一个重要内容就是城市居民的社会网分析。但由于社会网本身具有选择性,要分析社会网、社会资本对个人结果是否具有因果作用,就必须解决这个内生性问题。如果直接将社会网的属性(如熟人圈的平均社会地位、职业声望、教育程度等)对被访者的个人结果(如工资收入)进行简单回归,得到的估计量就必然有偏。通过考察上述(CGSS2003)问卷中有关“社会交往”和“获取当前工作”的题器设计,在借鉴前人研究和其他学科的方法的基础上¹,笔者提出两条有利于解决内生性问题的数据收集策略,期望在今后相关研究的资料收集与测量中得到使用与印证。

第一,提高社会网络数据收集层级。就CGSS2003的“社会交往”题器而言,数据收集仅局限于两层数据即被访者和被访者的朋友(题器运用提名法要求被访者提供不超过5个朋友或熟人的工作、教育、职业等信息)。在今后的设计中,建议设法对“朋友”的配偶、父母和朋友的朋友进行数据收集。这样就构成一个三层数据体系。也即被访者(第一层)-被访者的朋友(第二层)-被访者的朋友的其他社会关系(第三层)。这本质上就为“朋友”提供了一个工具变量,也即,第三层作为第二层的工具变量。该策略的提出理由是:第三层变量显然和第二层有关,但如果我们能确保第三层的朋友的社会关系和被访者不直接认识(譬

1. 相关的文献回溯详见:陈云松、范晓光:《社会资本的劳动力市场效应:关于内生性问题的文献回溯》,未刊稿,2010年。

如,被访者会听说自己的朋友的某些社会关系人的信息),那么第三层变量就会仅仅通过影响第二层变量来影响第一层变量。

第二,注意外生性数据的收集和挖掘使用。CGSS2003 问卷已经注意到了某些外生性变量的数据采集,如兄弟姐妹的数量、子女的性别和春节期间相互拜年交往的人数等,但诸如彩票号码、河流数目、政府外在政策干预等等都是外生的数据(它们都曾在相关的文献里被使用过),往往能够对于社会资本研究中内生性问题的解决起到作用。尤其是工具变量,它往往需要“灵感式”的逻辑推导,一些看似和研究无关的变量,反而就有可能成为解决内生性问题的奇兵。

三、结语

探求因果机制是包括社会学在内的社会科学所共同担负的学术责任。社会实验由于成本、研究伦理的限制,在短期内似乎较难成为社会科学研究的主导方式。因此,基于调查数据的定量分析,将长期是社会学家解释社会现象的重要工具之一。但所有从事定量分析的社会学家都应该认识到,在进行统计回归时我们所必须接受的假设以及这些假设对于因果判断特别是内生性问题的意义。所有基于非实验数据的定量分析研究都应该清楚地向读者说明,研究结论究竟是一般性的统计相关描述,还是在某个假设满足条件下的因果关系。而这些假设以及数据、模型识别策略对这些假设的满足程度,也都应该得到清晰的描述,以便读者有明确的判断,而不是含糊其辞。从社会互动研究乃至整个社会科学领域的研究脉络来看,尽管社会学家早于经济学家发现和提出了很多极富启发性且重要的社会现象和理论概念(例如格兰诺维特的嵌入性、林南的社会资源理论等),但在解决模型内生性问题并获得相对精准的因果判断方面,似乎没有经济学家们走得那么远。以运用工具变量著称的经济学家安格里斯特和克鲁格(Angrist & Krueger, 1999)就认为,劳动经济学已经走在社会科学领域的模型识别和解决内生性问题的前沿。

所幸的是,近年来社会学家对系统解决内生性问题逐渐开始重视,越来越多的社会学家在重新检视定量分析的应用和限度(Morgan & Winship, 2007; Mouw, 2006; 谢宇, 2006 & 2008)。一旦社会学家能够把提出概念、解析现象、构建理论的灵感和提高数据质量、完善模型设置

和详察因果机制有机结合起来,那么社会学对于整个社会科学领域的贡献必将获得新的突破。

最后,我们必须说明的是,和 T. 莫维(Mouw, 2006)一样,我们对社会学文献忽视内生性问题的批评以及对中国综合调查问卷设计的粗浅解读,绝非是对过去和当下的社会学研究进行“挑刺”,而是为了让更多的社会学研究来重视内生性问题,以对早期研究者的的重要理论贡献进行进一步的确认和证明。

参考文献(References)

- Aaronson, D. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes." *Journal of Human Resources* 33.
- Ammermüller, Andreas and Jörn-Steffen Pischke. 2006. "Peer Effects in European Primary Schools: Evidence from PIRLS." ZEW Discussion Paper No. 06 - 027, Mannheim.
- Angrist, Joshua D. and Alan B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics* III A., edited by Orley Ashenfelter and David Card, Amsterdam: North-Holland.
- Angrist, Joshua D. and Victor Lavy. [1999]2002. "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement." In *The Economics of Schooling and School Quality*, edited by R. Hanushek Edward Elgar, 2002.
- Araujo, Caridad. Alain de Janvry, and Elisabeth Sadoulet. 2004. "Peer Effects in Employment: Results from Mexico's Poor Rural Communities." University of California at Berkeley, Working Paper.
- Arcidiacono, Peter and Nicholson, Sean. 2005. "Peer effects in medical school." *Journal of Public Economics* 89(2 - 3).
- Asadullah, Mohammad Niaz and Chaudhury, Nazmul. 2008. "Social Interactions and Student Achievement in A Developing Country: An Instrumental Variables Approach." Policy Research Working Paper Series 4508, The World Bank.
- Bayer, P., Robert McMillan and Kim Rueben. 2002. "What Drives Racial Segregation? New Evidence Using Census Microdata." Working Papers. 02 - 26
- Bayer, P. and S. Ross. 2006. "Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application." NBER Working Paper 12211. Cambridge, United States: National Bureau of Economic Research.
- Bentolila, S., C. Michelacci, and J. Suarez. 2003. "Social Networks and Occupational Choice." London School of Economics, Centre for Economic Performance, mimeo.
- Berk, Richard A. and Subhash C. Ray. 1982. "Selection Biases in Sociological Data." *Social Science Research* 11.
- Berk, Richard A. and Subhash C. Ray. 1983. "An Introduction to Sample Selection Bias in Sociological Data" *American Sociological Review* 48.
- Betts, J. R. and A. Zau. 2004. "Peer Groups and Academic Achievement: Panel Evidence from Administrative Data." Unpublished manuscript.
- Blalock. 1984. "Contextual Effects Models: Theoretical and Methodological Issues." *Annual Review of Sociology* 10.
- Bobonis, Gustavo J. and Frederico Finan. 2008. "Neighborhood Peer Effects in Secondary

- School Enrollment Decisions.” Working paper.
- Boozer, Michael and Stephen E. Cacciola. 2001. “Inside the ‘Black Box’ of Project Star: Estimation of Peer Effects Using Experimental.” Working paper.
- Bramoullé, Y., H. Djebbari and Fortin B. 2007. “Identification of Peer Effects through Social Networks.” IZA Discussion Paper 2652. Bonn, Germany: Institute for the Study of Labor (IZA).
- Brock, William A. and Durlauf, Steven N. 2001. “Interactions-based Models.” In (ed.) *Handbook of Econometrics* (1st), edited by J. J. Heckman and E. E. Leamer Elsevier.
- Bushway, S. B. Johnson, and L. A. Slocum, 2007. “Is the Magic still There? The Relevance of the Heckman Two-step Correction for Selection Bias in Criminology.” *Journal of Quantitative Criminology* 23(2).
- Cameron and Trivedi. 2005. *Microeconometrics*. New York: Cambridge University Press.
- Card and Krueger 1996. “School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina.” *Journal of Economic Perspectives* 10.
- Carman, Katherine and Lei Zhang. 2008. “Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School.” Unpublished manuscript.
- Cason, Timothy N. and Vai-Lam Mui. 1998. “Social Influence in the Sequential Dictator Game.” *Journal of Mathematical Psychology* 42.
- 陈云松、范晓光. 2010. 社会资本的劳动力市场效应:关于内生性问题的文献回溯[R]. 工作论文. [Chen Yunsong Fan Xiaoguang. 2010. “The Labour Market Effect of Social Capital: A Literature Review of the Endogeneity Problem.” Working Paper.]
- 陈云松. 2008. 分析社会学:寻求连接微观与宏观的机制性解释[J]. 浙江社会科学(5).
- Chen Yunsong. 2008. “Analytical Sociology: Identifying the Mechanism-based Explanations Linking Micro and Macro.” *Zhejiang Social Science* (5). (in Chinese)]
- Cipollone, Piero and Rosolia, Alfonso. 2007. “Social Interactions in High School: Lessons from an Earthquake.” *American Economic Review* 97(3).
- Clampet-Lundquist, S. and Massey, D. 2008. “Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment.” *American Journal of Sociology* 114(1).
- Clark, Andrew E. and Youenn Lohéac. 2005. “‘It Wasn’t Me, It Was Them!’ Social Influence in Risky Behavior by Adolescents.” IZA Discussion Papers 1573, Institute for the Study of Labor (IZA).
- Conley, T. G. and C. R. Udry. 2008. “Learning About a New Technology: Pineapple in Ghana”(Sept 2008 Revision). *Forthcoming American Economic Review*.
- Contreras, Dante, Diana Kruger, Marcelo Ochoa and Daniela Zapata. 2007. “The Role of Social Networks in the Economic Opportunities of Bolivian Women.” Working paper.
- Currie, Janet and Yelowitz, Aaron. 2000. “Are Public Housing Projects Good for Kids?.” *Journal of Public Economics* 75(1).
- Cutler, David M. and Edward L. Glaeser. 1997. “Are Ghettos Good or Bad?.” *The Quarterly Journal of Economics* 112(3).
- Delattre and Sabatie. 2007. “Social Capital and Wages: An Econometric Evaluation of Social Networking’s Effects.” *Labour* 21 (2).
- DiPrete, Thomas A. and Markus Gangl. 2004. “Assessing bias in the estimation of causal effects: Rosenbaum bounds on Matching Estimates and Instrumental Variables with Imperfect Instruments.” *Sociological Methodology* 34.
- Duflo, Esther and Emmanuel Saez. 2003. “The Role of Information And Social Interactions In Retirement Plan Decisions: Evidence From A Randomized Experiment.” *The Quarterly*

- Journal of Economics* 118(3).
- Duncan, G. J. J. Brooks-Gunn and P. K. Klebanov. 1994. "Economic Deprivation and Early Childhood Development." *Child development* 65(2).
- Duncan, G. Connell, J. Klebanov, P. 1997. "Conceptual and methodological issues in estimating causal effects of neighbourhoods and family conditions on individual development." In *Neighbourhood Poverty: Context and Consequences for Children 1*, edited by J. Brooks-Gunn G. Duncan and J. Aber.
- Duncan, Greg J., Elizabeth Clark-Kauffman and Emily Snell. 2004. "Residential Mobility Interventions as Treatments for the Sequelae of Neighborhood Violence." Working paper, Northwestern University.
- Duncan, O. D., A. O. Haller and A. Portes. 1968. "Peer Influences on Aspirations: A Reinterpretation." *American Journal of Sociology* 74
- Durlauf and Fafchamps. 2004. "Social capital." NBER Working Paper 10485.
- Durlauf, Steven N. and Hisatoshi Tanaka. 2008. "Understanding Regression Versus Variance Tests For Social Interactions." *Economic Inquiry* 46(1).
- Durlauf. 2002. "On the Empirics of social capital." *Economic Journal, Royal Economic Society*. 112(483).
- Evans, W. N. W. E. Oates and R. M. Schwab. 1992. "Measuring Peer Effects: A Study of Teenage Behavior." *Journal of Political Economics* 100.
- Falk, Armin and Fischbacher, Urs. 2002. "'Crime' in the Lab-detecting Social Interaction." *European Economic Review* 46(4-5).
- Galbiati, Roberto and Giulio Zanella. 2008. "The Social Multiplier of Tax Evasion: Evidence from Italian Audit Data." Department of Economics, University of Siena 539.
- Glaeser, Edward L., Bruce Sacerdote and Scheinkman, Jose A. 1996. "Crime and Social Interactions." *The Quarterly Journal of Economics* 111(2).
- Gould, Eric D., Victor Lavy and M. Daniele Paserman. 2005. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." IZA Discussion Papers 1883.
- Graham, 2008. "Identifying Social Interactions Through Conditional Variance Restrictions." *Econometrica* 76(3).
- Grogger, J. 1996. "School Expenditures and Post-schooling Earnings: Evidence from the High School and Beyond." *Review of Economics and Statistics* 78.
- Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30.
- Hanushek, E., Kain JF, Markman JM and Rivkin SG. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrica* 18(5).
- Hanushek E, S. Rivkin and Taylor, L. 1996. "Aggregation and the Estimated Effects of School Resources." *Review of Economics and Statistics* 78.
- Harding, D. J. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy." *American Journal of Sociology* 109.
- Hartmann, W. R. et al. 2008. "Modeling Social Interactions: Identification, Empirical Methods, and Policy Implications." *Marketing Letters* 19.
- Hartmann, Wesley and Taylan Yildiz. 2006. "Preferences for Joint Consumption and the Implications for Targeted Marketing." Working paper.
- Heckman, J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and A Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5.

- Hedström and Swedberg. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Hedström. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Hedström. 2008. "Studying mechanisms to strengthen causal inferences in quantitative research." in *The Oxford Handbook of Political Methodology*, edited by J. M. Box-Steffensmeier, H. E. Brady and D. Collier. New York: Oxford University Press.
- Holland, P. W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81.
- Hoxby and Weingarth. 2005. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Working Paper, Harvard University.
- Hoxby, CM. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working paper. 7867.
- Hughes, M. R. 1997. "Sample Selection Bias in Analyses of the Political Democracy and Income Inequality Relationship." *Social Forces* 75(3).
- Ichino, Andrea and Giovanni Maggi. 2000. "Work Environment And Individual Background: Explaining Regional Shirking Differentials In A Large Italian Firm." *The Quarterly Journal of Economics* 115(3).
- Jacob, B. 2004. "Public Housing, Housing Vouchers & Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94(1).
- Jencks, C. Mayer, S. E. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." In *Inner-City Poverty in the United States*, edited by L. E. Lynn and M. G. H. McGeary. Washington, DC: National Academy Press.
- Kandel, Denise B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology* 84.
- Katz, L. , J. Kling and J. Liebman. 2001. "A Moving to Opportunity in Boston: Early Results of A Randomized Mobility Experiment." *Quarterly Journal of Economics* 116.
- Kawaguchi, Daiji. 2004. "Peer Effects on Substance Use among American Teenagers." *Journal of Population Economics* 17(2).
- Kling, Jeffrey, J. B. Liebman, Lawrence F. Katz and Lisa Sanbonmatsu. 2004. "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment." mimeographed, Princeton University.
- Kling, Jeffrey R. , Jeffrey B. Liebman and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75.
- Kling, Jeffrey R. , Jens Ludwig and Lawrence F. Katz. 2004. "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Mobility Experiment." *Quarterly Journal of Economics* 120.
- Krauth, Brian. 2004. "Peer Effects and Selection Effects on Smoking among Canadian Youth." *Canadian Journal of Economics* 38(3).
- Lavy, Victor, Daniel Paserman and Analia Schlosser. 2008. "Inside the Black Box of Ability Peer Effect: Evidence from Variation of Low Achiever in the Classroom." NBER working paper No 14415.
- Liebman, J. , L. F. Katz and J. R. Kling. 2004. "Beyond Treatment Effects: Estimating the Relationship between Neighborhood Poverty and Individual Outcomes in the MTO Experiment." Princeton University Industrial Relations Section Working Paper No. 493.
- Lin and Ao, 2008. "The Invisible Hand of Social Capital: An Exploratory Study." In Lin & Ericson, 2008, *Social Capital: An International Research Program*, New York: Oxford

University Press.

- Ludwig, J., G. Duncan and J. C. Pinkston. 2005. "Housing Mobility Programs and Economic Selfsufficiency: Evidence from A Randomized Experiment." *Journal of Public Economics* 89(1).
- Manski, C. 1993. "Identification of Endogenous Social Effects: the Reflection Problem." *Review of Economic Studies* 60.
- Manski, C. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14(3).
- Marsden, Peter and Jeanne Hurlbert. 1988. "Social Resources and Mobility Outcomes: A Replication and Extension." *Social Forces* 66.
- Marsden, Peter and Jeanne Hurlbert. 2001. "Interpersonal Ties, Social Capital, and Employer Staffing Practices." in *Social Capital: Theory and Research*, edited by N. Lin, K. S. Cook and R. Burt. Hawthorne, NY: Aldine de Gruyter.
- Maurin, E. and J. Moschion. 2009. "The Social Multiplier and Labor Market Participation of Mothers." *American Economic Journal: Applied Economics* 1(1).
- McEwan, Patrick J. 2003. "Peer Effects on Student Achievement: Evidence from Chile." *Economics of Education Review* 22(2).
- McPherson, Miller, Lynn Smith-Lovin and James Cook. 2001. "Birds of A Feather: Homophily in Social Networks." *Annual Review of Sociology* 27.
- Moffit, R. 1995. "Selection Bias Adjustment in Treatment-effects Models as a Method of Aggregation." *Proceedings of the American Statistical Association*.
- Moffit, R. 2001. "Policy Interventions, Low-Level Equilibria and Social Interactions." In S. N. Durlauf, P. H. Young (ed.) *Social Dynamic*. Boston, MA: MIT Press.
- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Morgan, Stephen L. 2001. "Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on Learning." *Sociology of Education* 74.
- Morgan, Stephen L. and David J. Harding, D. 2006. "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice." *Sociological Methods and Research* 35.
- Mouw, Ted. 2006. "Estimating the Causal Effect of Social Capital: A Review of Recent Research." *Annual Review of Sociology* 32.
- Mouw, Ted. 2003. "Social Capital and Finding a Job: Do Contacts Matter?" *American Sociological Review* 68.
- Nair, Harikesh, Puneet Manchanda and Tulikaa Bhatia. 2006. "Asymmetric Peer Effects in Physician Prescription Behavior: The Role of Opinion Leaders." SSRN working paper.
- Nan Lin. 2004. "Job Search in Urban China: Gender, Network Chains, and Embedded Resources." in *Creation and Returns of Social Capital: A New Research Program*, edited by Henk Flap and Beate Volker London and New York: Routledge.
- Palloni A, D. Massey, M. Ceballos, K. Espinosa and M. Spittel. 2001. "Social Capital and International Migration: A Test Using Information on Family Networks." *American Journal of sociology* 106
- 波普诺,戴维. [1995]1999. 社会学(第十版)[M]. 北京:中国人民大学出版社. [Popenoe David (1995)1999. *Sociology*(Tenth Edition). Beijing:China Renmin University Press.]
- Rivkin, Steven. Tiebout Sorting. 2001. "Aggregation and the Estimation of Peer Group Effects." *Economics of Education Review* 20(3).
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70.
- Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results with Dartmouth

- Roommates." *Q. J. Econ.* 116(2).
- Sampson, Robert J. 2008. "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure." *American Journal of Sociology* 114 (1).
- Sanbonmatsu L, J. R. Kling, G. J. Duncan and Gunn J. Brooks. 2006. "Neighborhoods and Academic Achievement: Results From the Moving to Opportunity Experiment." NBER Working Paper 11909.
- Sobel, Michael E. , 1996. "An Introduction to Causal Inference." *Sociological Methods and Research* 24.
- Soetevent, A and Peter Kooreman. 2005. "Social Ties within School Classes: The Roles of Gender, Ethnicity, and Having Older Siblings." *Oxford Review of Economic Policy* 21(3).
- Soetevent, A. 2006. "Empirics of the Identification of Social Interactions: An Evaluation of the Approaches and Their Results." *Journal of Economic Surveys* 20(2).
- Sorensen, A. 2005. "Social Learning and Health Plan Choice." Forthcoming, RAND Journal of Economics.
- Stolzenberg, R. M. and D. A. Relies. 1997. "Tools for Intuition about Sample Se-lection Bias and its Correction." *American Sociological Review* 62.
- van der Klaauw and van Ours, 2003. "From Welfare to Work: Does the Neighborhood Matter?" Working paper.
- Vigdor, Jacob and Thomas Nechyba. 2004. "Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment." unpublished manuscript.
- 王天夫. 2006. 社会研究中的因果分析[J]. 社会学研究(4). [Wang Tianfu. 2006. "Causal Analysis in Social Research." *Sociological Research*(4). (in Chinese)]
- Wegener, B. 1991. "Job Mobility and Social Ties: Social Resources, Prior Job, and Status Attainment." *American Sociological Review* 56.
- Winship, C and S. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25: 659 - 706.
- Winship, C and Sobel, M. 2004. "Causal Inference in Sociological Studies." In *Handbook of Data Analysis, A Bryman*, edited by M. Hardy. Thousand Oaks, CA: Sage.
- Wooldridge, J M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Wooldridge, J M. 2006. *Introductory Econometrics: A Modern Approach* (3rd Edition) Chapters, Thomson South-Western.
- 谢宇. 2006. 社会学方法与定量研究[M]. 北京: 社会科学文献出版社. [Xie Yu. 2006. *Sociological Methods and Quantitative Research*. Beijing: Social Sciences Academic Press.]
- 谢宇. 2008. 奥迪斯·邓肯的学术成就: 社会科学中用于定量推理的人口学方法[J]. 社会(3). [Xie Yu. 2008. "Otis Dudley Duncan's Legacy: the Demographic Approach to Quantitative Reasoning in Social Science." *Society*(3). (in Chinese)]
- 章奇. 2008. 社会科学中的因果关系及其分析方法[J]. 浙江社会科学(3). [Zhang Qi. 2008. "Causation and Causal Analysis in Social Sciences." *Zhejiang Social Science*(3). (in Chinese)]
- Zimmerman, D. 2003. "Peer Effects in Academic Outcomes: Evidence From a Natural Experiment." *Review of Economics and Statistics* 85(1).

责任编辑: 路英浩