

逻辑、想象和诠释: 工具变量 在社会科学因果推断中的应用*

陈云松

提要: 工具变量(instrumental variable) 是社会科学定量分析中解决内生性问题的重要手段, 是基于调查数据进行因果推断的前沿方法。本文在简要介绍工具变量的定义、原理及估算方法的基础上, 对实证分析中较为常见的五类工具变量进行回顾梳理, 为今后研究寻找工具变量提供了参考。同时, 对工具变量估计量的权重性特征进行了阐述, 并结合实例展示了使用工具变量进行因果推断的基本步骤和要点。最后, 就工具变量方法的潜力和局限性进行了剖析。

关键词: 工具变量 内生性 定量分析 因果推断

一、导言: 因果推断的圣杯^①

在反事实因果的框架之下,^② 基于调查数据的社会学定量分析要进行因果推断, 难度极大。其主要原因在于, 社会学家一旦要证明某个他们所感兴趣的“因”会带来一定的“果”, 就必须面对一个永恒挑战: “内生性”问题(endogeneity)。也即: 如果某个潜在的、无法观测的干

* 作者感谢康奈尔大学摩根(Stephen Morgan)教授、牛津大学斯奈德斯(Tom Snijders)教授和赫斯特罗姆(Peter Hedström)教授、香港科技大学吴晓刚教授、中山大学梁玉成副教授、浙江省社会科学院范晓光助理研究员的讨论、批评和建议, 同时感谢匿名评审专家和编辑部提出的修改意见以及 CGSS 2003 和 CHIPS 2002 数据团队。文责自负。

① 圣杯(grail) 系指耶稣受难时用来盛放鲜血的圣餐杯。在凯尔特神话中, 追寻圣杯是一个神圣而伟大的主题, “有能者居之”, 其过程则充满艰险, 传说有无数骑士为了寻求圣杯而踏上了不归之路。

② 关于社会科学定量分析中的反事实因果框架可参见摩根和温肖普的文章(Morgan & Winship, 2007)。谢宇(2006)曾经以大学教育为例来说明反事实因果的含义。比如, 在分析大学教育对个体的收入是否有因果效应时, 对于一个上大学的学生, 我们不可能获得他不上大学情况下的数据。因此, 社会学定量分析中只能用平均干预效应来替代, 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。

扰项,既影响“因”,又影响“果”,那么,利用最小二乘法模型(简称 OLS 模型)进行回归分析所得到的估计量就会是有偏误的,而不具有因果推断力。在实证分析中,无论是经典的教育回报研究(Card, 1999),还是我国学界非常关注的关系网、社会资本研究(Mouw, 2003, 2006; 陈云松、范晓光, 2010, 2011),内生性问题都极为重要且亟待解决。

解决内生性问题的常见方法,主要包括工具变量(instrumental variable, 简称 IV)、固定效应模型(fixed effects model, 简称 FE)、倾向值匹配(propensity score matching, 简称 PSM)、实验以及准实验(experiments and quasi-experiments)等等。近年来,其中不少方法已经逐步在我国社会学界得到评述和应用(梁玉成, 2010; 陈云松、范晓光, 2010, 2011; 陈云松, 2012; 胡安宁, 2012; 魏万青, 2012)。^①在反事实因果分析框架下,实验或准实验方法最切近要义。但社会科学的很多研究主题和领域决定了无法使用实验方法,而其他方法也都具有较大的局限性。如固定效应模型只能消除时间固定的干扰项,倾向值匹配方法则完全依赖于“可观测因素被忽略”的假说。相比较而言,对基于调查数据的定量分析,工具变量方法具有独特优势。不过,工具变量方法在社会学分析中的运用,目前却远远不如它在计量经济学和政治学定量分析中那么广受青睐(政治学和计量经济学研究中工具变量方法的使用参见 Sovey & Green, 2011; Angrist et al., 1996; Angrist & Krueger, 2001)。

但是,近 10 年来社会学界对于工具变量的态度,正在由不熟悉、犹豫不决向着逐步接纳而转型。10 年前,康奈尔大学的摩根有感于工具变量方法得不到社会学家青睐而专门写成一篇文章《社会学家该不该用工具变量》(Morgan, 2002)。整整 10 年后,美国《社会学年鉴》(*Annual Review of Sociology*)专门刊出了博伦关于工具变量在社会学分析中应用的综述(Bollen, 2012)。这篇重要论文从技术角度详细回顾了 2000 - 2009 年间在美国三大顶级社会学刊物(*American Sociological Review*, *American Journal of Sociology*, 以及 *Social Forces*) 刊发的 57 篇采用工具变量方法的论文。毫无疑问,工具变量方法逐步被社会学界关注和接纳的过程,充分展示了社会学定量分析方法的演进以及与其他

① 在社会资本和社会网研究中对内生性问题关注最早的仍是计量经济学界(参见 Manski, 1993, 2000; Moffitt, 2001; Durlauf, 2002; Durlauf & Fafchamps, 2004; Bramoullé et al., 2007)。

学科在方法论上的进一步融合。

那么,何以说工具变量是定量分析中因果推断的“圣杯”?这是因为好的工具变量非常难以寻觅,寻找它的逻辑和数据挖掘过程充满艰辛、难以驾驭,甚至往往需要研究者的灵感。但它在模型上的简洁性,它对社会科学想象力、逻辑力和诠释力的要求,既为定量分析提供了因果推断的重要武器,也让分析的过程充满趣味和奇思妙想。本文将以尽量浅显的语言,总结回顾工具变量的原理、来源、分类和特性,并结合具体分析案例,勾勒工具变量方法的实施步骤和诠释要点。本文是当前社会学文献中首次对工具变量进行梳理和分类的尝试,也是目前我国社会学定量分析领域第一次对工具变量的全面介绍、评述和案例展示。

二、工具变量的原理:模型之外的力量

工具变量的原理最早由菲利普·莱特(Philip G. Wright)在上世纪20年代末提出(Stock & Trebbi 2003),这里仅作扼要介绍和基本的模型推演。首先,我们给出一个典型的线性回归模型:

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

这里 y 为因变量,也即“果”; x_1 为自变量,或者解释变量,也即“因”。大写的 X 为外生控制项向量(也即一组假定为外生的其他控制变量,例如年龄、性别等等), ε 则为误差项。如果 ε 与 x_1 不相关,那么我们可以利用 OLS 模型对方程进行无偏估计。然而,如果一个重要变量 x_2 被模型(1)遗漏了,且 x_1 和 x_2 也相关,那么对 β_1 的 OLS 估计值就必然是有偏的。此时 x_1 被称作“内生”的解释变量,这也就是著名的“内生性”问题。

要解决这一内生性问题,我们需要引入更多信息来进行无偏估计。工具变量的方法就是引入一个外生变量 Z ,且 Z 必须满足以下两个条件:与 ε 不相关,但与 x_1 相关。或者说, Z 仅仅通过影响 x_1 来影响 y 。这样,根据工具变量的必备条件,我们可以得到:

$$\text{Cov}(Z, x_1) \neq 0; \text{Cov}(Z, \varepsilon) = 0 \quad (2)$$

由方程(1)我们可以推导出:

$$\text{Cov}(Z, y) = \beta_1 \text{Cov}(Z, x_1) + \beta \text{Cov}(Z, X) + \text{Cov}(Z, \varepsilon)。$$

再根据方程(2)和 X 是外生向量的假设,我们得到 $\text{Cov}(Z, y)$

$= \beta_1 \text{Cov}(Z, x_1)$,也即:

$$\beta_1 = \text{Cov}(Z, y) / \text{Cov}(Z, x_1)$$

故此, 我们可以对 β_1 进行无偏估计:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})(y_i - \bar{y})}{\sum_{i=1}^n (Z_i - \bar{Z})(x_{1i} - \bar{x}_1)} \quad (3)$$

方程(3)里的 $\hat{\beta}_1$,也就是工具变量估计量。

如果用上述公式还不能直观清晰地表达工具变量的原理, 那么我们可以用下面的示意图来做一简要说明。在图1中, 模型的范围用虚线框来表示。工具变量 Z 处于模型之外(也即在虚线框之外), 因此是完全外生的。^① 此时, 工具变量 Z 只能通过影响自变量 x_1 而间接影响因变量 y 。如果工具变量 Z 和自变量 x_1 密切相关, 那么, 只要工具变量 Z 有了增量变化, 就必然会对自变量 x_1 产生一个来自模型之外的冲击。如果自变量 x_1 和因变量 y 之间真的存在因果关系, 那么 Z 对 x_1 带来的冲击也就势必传递到 y 。这样, 在一系列的假说之下, 只要 Z 对 y 的间接冲击能够被统计证明是显著的, 我们就可以推断出 x_1 对 y 必然有因果关系。利用对 Z 与 x_1 相关的估算, 以及 Z 与 y 的间接相关的估算, 理论上我们就可以推导出 x_1 和 y 之间真实关系的大小 β_1 。因此, 图1非常清晰地展示了工具变量的原理: 利用来自模型之外的外生差异进行无偏估计。

较为常见的工具变量估算方法是两阶段最小二乘法 (two-stage least-squares, 也即 2SLS)。在回归的第一阶段, 内生的因变量 x_1 放在模型左侧, 而右侧则为原模型中全部 X 以及工具变量 Z 。然后对每一个 x_1 进行预测赋值。在第二阶段, 模型左侧是因变量 y , 右侧则为 X 和 x_1 的第一阶段预测值。工具变量估计量肯定是一致的(参数估计的一致性指当样本容量趋向无穷大时, 参数估计趋近于参数真值, 也就是收敛于参数真值)。不过其估计方差也比相应的 OLS 估计方差要大。工具变量 Z 和自变量 x_1 之间的关系越紧密, 则估计方差越小。

要确保工具变量分析结果稳健可信, 我们必须首先检验工具变量的合法性, 同时还要观察工具变量模型和一般的单方程模型(如 OLS 或 Probit 模型)之间的分析结果有无系统差异。这里, 有几个至关重要

① 当然, 模型内的干扰项和主解释变量以及因变量之间的关系也可能是逆向的。

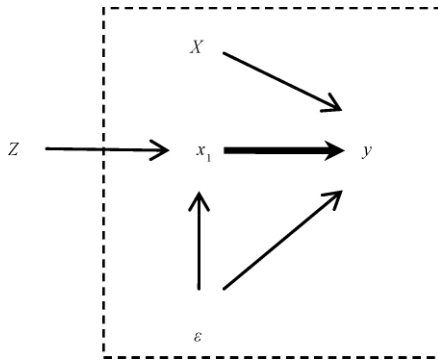


图1 模型之外的力量: 工具变量原理

的统计量的判别是必不可少的。(1) Z 和 x_1 必须是强相关。否则, 就会带来弱工具变量问题, 导致估计量有偏。在使用二阶段估计法时 (2SLS), 一般我们可以依赖 F 统计量来判断两者之间的关系强弱。一般而言, 如果 F 统计量大于经验值 10, 则不存在弱工具变量问题 (Stock & Yogo, 2005)。(2) 工具变量的外生性是无法用统计方法直接验证的。不过当我们同时使用多个工具变量时 (也即模型被过度识别时 overidentified), 则可以进行沙根检验 (Sargan Test); 此外, 有一些研究会把工具变量直接加入主模型进行偏系数的显著性检验, 不过严格意义上这一做法并不具有有效性。(3) 豪斯曼内生性检验 (Hausman Test of Endogeneity), 用以检测 OLS 模型和工具变量模型之间是否存在系统差异。如果有, 则应采纳工具变量估计量。(4) 瓦尔德内生性检验 (Wald Test of Endogeneity), 用以检测单方程 Probit 模型与 IV-Probit 模型之间是否存在系统差异。如果有, 则应采纳 IV-Probit 模型估计量。

三、工具变量的寻觅: 逻辑和想象

本部分将对一些经典和较具说服力的工具变量分析案例进行分类总结。梳理这些工具变量, 比一般性的文献综述更重要, 这是因为, 合格的工具变量非常难以寻找。因此, 前人对某一类工具变量的使用, 在很大程度上对我们今后寻找工具变量能够带来重要启发甚至灵感: 严

密的逻辑和辽远的想象力,是寻找到好的工具变量的必要条件。^①

(一) 来自“分析上层”的工具变量: 集聚数据

经济学和社会学中一个非常热门的研究课题是同侪效应(peer effect)。其假说是,个人的经济社会结果,往往会受到所在集体的某个特征要素的影响。比如,一个人的成绩、收入、社会地位等等,会受到他所在的学校、班级、邻里的特征的影响。但要验证这一假说,我们就必须解决个人异质性导致的内生性问题。这是因为,很多无法观测到的个人、家庭因素,会同时和个人结果与我们关心的集体要素相关(特别是,个体往往根据自己的某项特质和偏好来选择学校、选择班级、选择邻居)。为解决这一内生性问题,经济学家和社会学家常常把州、县或大都会地区层面的集聚数据(aggregation data)作为学校、班级和邻里等层面解释变量的工具变量(文献回溯参见 Card & Krueger, 1996)。例如,埃文斯等(Evans et al., 1992)试图验证学校中的贫困生比例对学生怀孕或辍学行为是否有显著影响。他们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。其理由是:以都会为单位的失业率和贫困率必然和辖区内学校的贫困生比例有关,但又不直接影响学生的怀孕或辍学等行为。邦托利阿等(Bentolila et al., 2010)使用联邦就业率作为工具变量来分析“使用社会关系”对个人收入是否具有作用。联邦就业率与收入没有直接关系。但就业率高,则在联邦内使用关系求职的必要性就低。不过,使用集聚层数据作为工具变量,往往会引入噪音,甚至增加遗漏偏误——因为我们无法保证高级区划层面上的特征值是完全外生的(Grogger, 1996; Hanushek et al., 1996; Moffitt, 1995; Rivkin, 2001)。因此,这类工具变量的应用案例,在近年来的研究中有减少的趋势。

(二) 来自“自然界”的工具变量: 物候天象

河流、地震、降雨、自然灾害等自然现象在一定地域范围内具有高度的随机、外生特性,因此可以被假设为与个人和群体的异质性无关,

^① 在面板数据分析中,一个标准化的方法是:解释变量的若干期滞后的线性组合(lagged terms)可以用来做工具变量。但这需要一定的假设。限于篇幅,且该方法不需要“寻找”工具变量,本文不做介绍。

同时,它们又能够影响一些社会过程。例如,霍克斯比(Hoxby,2000)在一篇经典研究中,采用区域内河流数量作为该区域学校数量的工具变量,以此来验证学区内的学校竞争是否可以提高教学质量。区域内学校数量之所以是内生的,是因为它可能是该区域长期历史积累下的某种特征的结果。而使用河流数作为工具变量则具有很强的说服力:河流数量越多,就会因交通问题导致更多学校的设立;但河流数是天然形成的,本身和教学质量无直接关系。再如,卡尔特和克莱泽(Culter & Glaeser,1997)把贯穿大都市的河流数量作为邻里区隔(segregation)的工具变量,以分析区隔程度对居住者的影响。该工具变量的合法性与上例工具变量异曲同工:河流越多,导致的邻里区隔程度必然越大,而河流数量和作为社会结果的居住者的收入无关。

除了河流,其他诸如地震、灾害、降雨量甚至化学污染等自然现象,都曾被具有社会科学想象力的研究者所使用。例如,在班级效应(class effect)研究中,西波隆和罗索利亚(Cipollone & Rosolia,2007)以地震导致的男性免征兵政策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成绩的影响。地震作为一种天象,显然是随机和外生的。在国际移民研究中,孟希(Munshi,2003)则使用墨西哥移民来源地区的降水量作为移民数量的工具变量,证明了同乡的移民越多,他们在美国打工的收入会越高。移民来源社区的降雨量作为工具变量的理由是:墨西哥某社区的降水量和美国的劳动力市场显然没有任何关联;但降水量和社区的农业收入有关,并通过影响农业预期收入而影响到移民美国的决策。在社会资本和网络效应研究中,陈云松(2012)以中国农民工来源村庄的自然灾害强度作为本村外出打工者数量的工具变量,证明了同村打工网的规模直接影响农民工在城市的收入。使用自然灾害作为工具变量的合法性在于:灾害越重,外出打工的村民就越多;而在控制了地区间应对灾害的能力和来源省份之后,发生在村庄领域内的自然灾害可以被认为是外生的。

(三) 来自“生理现象”的工具变量: 生老病死

人类的生老病死既是社会现象,也是生理上的自然现象。出生日期、季度、性别、死亡率等,虽仅仅是有机体的自然历程,但既具有随机性,又往往和特定的经济社会过程相关。因此,无论在宏观还是微观社会科学层面,它们都曾被巧妙地作为工具变量运用在因果推断之中。

比如,在制度经济学研究中,我们关心的是制度对一国的人均收入有无影响。但制度往往是内生的,或者说是选择性的(例如,好的制度也许总在人均收入高的国家或地区产生)。只有找到制度的工具变量,才能让人信服地证明制度的力量。阿西莫格鲁等在一项经典研究(Acemoglu et al. 2001)中,把殖民地时代一个国家的自然死亡率作为该国当今制度的工具变量。其理由非常巧妙:如果该地区当年的死亡率高,那么欧洲殖民者就相对不愿定居下来,从而在当地建立起更具掠夺性的“坏”制度。由于制度的“路径依赖”,殖民时代的制度显然和现在的制度关系密切。因此,历史上的死亡率作为工具变量,应该和当今制度紧密相关,而一百年前的死亡率作为一种自然生理现象,又和目前的人均收入没有直接关系。

在微观层面的研究中,个人的出生时段曾多次被作为工具变量使用。在教育回报研究中,安古瑞斯特和克鲁格(Angrist & Krueger, 1991)把被访者出生的季度作为教育的工具变量。其理由是:上半年出生的孩子退学的可能性大于下半年出生的孩子(美国《义务教育法》规定不满16周岁不得退学),因此后者平均受教育时间更长。除了个体的“生日”,人类的生育结果也往往作为一种随机现象而被当作工具变量使用。例如,安古瑞斯特和伊凡斯(Angrist & Evans, 1998)试图分析家庭中的孩子数是否影响母亲的就业。由于生育孩子数量是可以被选择的,因此解释变量显然是内生的。为解决这一问题,他们巧妙地挖掘了人类生育行为中偏好有儿有女的特征,将子女“老大”和“老二”的性别组合情况作为工具变量。理由是:头两胎如果是双子或双女,那么生育第三胎的可能性大大增加,进而增加子女数。而子女性别是完全随机的。莫林和莫斯基翁(Maurin & Moschion, 2009)考察了法国邻里中其他母亲的就业如何影响单个母亲的就业。为了消解内生性问题,他们用邻里平均的头两个子女的性别组合作为邻里母亲就业的工具变量。理由是,邻里平均的头两个子女的性别组合,会影响到邻里内母亲的平均就业情况,而邻里的下一代性别结构是随机的,和单个母亲的就业决定没有直接关系。前文提及的邦托利阿等人(Bentolila et al., 2010)的研究,在使用联邦就业率作为“使用社会关系”的工具变量的同时,还使用年长的兄姊数目作为工具变量。他们的理由是,兄弟姐妹数是随机的自然现象,因为家庭育儿数字在法国是随机的。但兄弟姐妹越多,则社会关系越多,托人帮助求职的可能性也就越大。

(四) 来自“社会空间”的工具变量: 距离和价格

社会空间的载体,包括具象性的城市、乡村和非具象性的市场空间等,和人类的行为与社会结果息息相关,但往往又在特定分析层面上具有独立性、随机性。这类经典的工具变量分析案例,出现在教育回报研究中。卡德(Card,1995)使用被访者的家到最近的大学的距离作为教育的工具变量,以此来分析教育是否能增加个人的收入和地位。我们知道,教育作为解释变量之所以是内生的,是因为人们会选择上或不上大学。而从家到大学的距离,会影响到是否上大学这个理性选择。但这一距离,作为城市空间的要素,显然又与个体的社会经济结果没有直接关系。当然,地理空间上的距离对于上学意愿的影响可能是微弱的,或者只影响到一部分人。这也就引出工具变量研究中的两个重要概念——弱工具变量和局部干预效应问题。这些在本文的下节将做说明。

在制度分析的研究中,霍尔和琼斯(Hall & Jones,1999)则非常具有想象力地用各国到赤道的距离作为工具变量。理由是,到赤道的距离大致反映了各国受西方制度影响的深浅程度,而这一距离显然是外生的。此外,除了城市距离这样具体的社会空间要素,市场作为社会经济活动的空间,其要素也往往和社会学家关心的现象紧密相关,却又不直接干扰个体的某些具体社会特征。钱楠筠(Qian,2008)巧妙地用茶叶的价格作为中国家庭中男性收入和女性收入之比的工具变量,证明了家庭收入的性别结构最终影响了中国家庭男女出生性别比例(传统上我们相信家庭的总收入影响男女出生比例)。用茶叶价格作为工具变量,是因为茶叶产业链的特性决定了从业人员以女性为主,茶叶价格提高就意味着女性在家庭中的经济地位提高,从而导致家庭女性胎儿被流产的几率降低。而茶叶的价格显然和家庭男女出生性别比例没有其他任何的因果逻辑联系。

(五) 来自“实验”的工具变量: 自然实验和虚拟实验

实验是一种外来的人为干预。它一方面对我们关心的解释变量带来冲击,同时又会置身模型之外。能够给我们提供工具变量的实验,既有政策干预、改革创新这样的社会实验,又有假想的虚拟实验。不少研究采用外生性政策干预所带来的自然实验来挖掘适当的工具变量。这方面,工具变量的重要权威安古瑞斯特有着极为经典的越南老兵系列研究。在越南战争期间,美国青年服兵役采取基于生日的抽签形式来

决定。获得抽签号之后,小于一定“阈值”就去参加体检服兵役,大于阈值的则可免于兵役。抽签号的产生如同彩票一样是完全随机的,但又直接影响到是否服兵役这一重要的社会过程。利用各人获得的抽签号作为工作变量,安古瑞斯特等人从80年代末到今天的20多年间,做出了一系列经典研究,分析服兵役是否会对当年的参战者、今天退伍老兵产生在收入、后续教育及健康等方面的影响(参见 Angrist, 1990, 1993; Angrist et al., 1996; Angrist et al., 2010)。^① 因为服兵役必然会减少教育年限,安古瑞斯特甚至尝试把抽签号作为教育的工具变量以分析教育回报(Angrist & Krueger, 1992)。^②

在同侪效应研究中也可以使用自然实验方法。例如,布泽和卡拉(Boozer & Cacciola, 2001)试图证明班级平均成绩对个体学业成绩具有同侪效应。他们把班级中曾经参与过“小班实验”的人数比例作为班级平均成绩的工具变量。其理由是:小班实验是随机设计的,学校之前从各个班级随机抽人去组成小班,因此外生性可以保证。而小班教学提高了这部分学生的成绩,故此对班级整体也必然有影响。类似方法还可参见博沃尔斯和菲南(Bobonis & Finan, 2008)对墨西哥教育的研究。他们把村庄中参与 Progresa 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。实际上,这一系列的工具变量思路来自于早前安古瑞斯特和拉维关于以色列班级规模效应的经典研究(Angrist & Lavy, 1999):小班是否有利于学生提高成绩。他们采用的工具变量是所谓迈蒙尼德(Maimonides)规则下的本地招生人数。根据迈蒙尼德规则,凡是班级人数超过40的就会被分裂为两个小班。因此,招生人数和班级规模之间就存在一个紧密的非线性关系。

更宏观社会空间层次的“自然实验”则是历史过程或大规模的社会运动。方颖、赵扬(2011)为估算各地区产权保护制度对经济增长的贡献,采用的工具变量是1919年中国不同城市基督教初级教会小学注册人数在当地人口中的比例。因为这一比例代表了该地区历史上受西方影响的程度:入读教会小学的人口比例越大,该地区受西方影响越

-
- ① 该工具变量后来引发了弱工具变量的争论。此外,抽签当兵者可能会更加关注健康,从而使得工具变量和因变量直接相关,导致工具变量估计量有偏。
 - ② 不过,该研究引起了较大争议。因为抽签号大于阈值的人会钻政策的空子而采用接受教育来逃避战争,由此造成抽签号不再外生。

大,也就越有可能在今天建立起较好的产权保护制度。而教会小学建立的初衷在于布道,这一变量和当年以及现在各地区的经济水平并不直接相关。在社会网分析中,章元、陆铭(2009)在分析农民工的家庭网、亲友网和收入之间的关系时,用农民工祖辈的社会背景及是否来自革命老区作为工具变量。其理由是,祖辈社会背景和是否来自老区会影响到农民工的社会网络规模,但这些历史因素和今天农民工在异地的收入没有其他直接联系。

虚拟实验则是一种非常有意思的工具变量方法,其原理有点接近于匹配方法。巴耶尔和罗斯(Bayer & Ross, 2009)在研究就业是否存在邻里同侪效应时,为解决被访者个体异质性和自选择问题,他们用与被访者具有相同个人特征(也即方程1中的 X 中的一组控制变量)的其他被访者的平均邻里特征来作为被访者邻里特征的工具变量。其理由是,相似个体选择相似的邻里,那么基于类似者的平均邻里特征(也即工具变量)应该与被访者的邻里特征相关。而同时他们利用了控制变量是外生的这一标准假设:既然个体的可观测特征与未被观察因子无关(也即方程中的 X 事先假定是外生的),那么基于 X 的一组外生变量而生成的工具变量也就必然与个人异质性无关。孔特雷拉斯等(Contreras et al., 2007)采取类似的方法,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。

四、工具变量估计量的诠释:局部干预效应问题

把模型外的变量 Z 引入模型时,既为模型的识别增加了有效信息,但同时也带来了噪音。其中一个最为重要的问题是:我们并不知道工具变量和解释变量之间相关的具体形式和特征。特别是,工具变量和解释变量之间的相关性,在样本中是均匀还是不均匀分布的呢?如果工具变量对样本人群的影响分布并不均匀,那么工具变量估计量应该更多地反映了部分样本上的因果效应。这就引出了工具变量分析中一个非常重要的解释框架——所谓的“局部平均干预效应”问题(local average treatment

effect 简称为 LATE。参见 Imbens & Angrist ,1994; Angrist et al. ,1996)。^① 这一问题之所以重要,是因为社会也好、区域也好、城市也好、人群也好,群体往往具有内在的差异性。这种差异性导致样本对于外生冲击的反应步调并不一致。这样,工具变量估计量在很多情况下只是基于诸多局部样本估计量的加权平均值。理解了这一权重特性,我们就能更好地对工具变量的分析结果进行合理的社会学解释,而不仅仅是提供一堆表格和纯技术性的统计量。

本文以前面提及的陈云松(2012)为例来做个说明。该研究试图证明农民工收入和同村打工网规模之间存在因果效应,数据来自 CHIPS 2002 共 22 个省的农户调查。作者把赫克曼二阶段模型(此处简称为 Heckit 模型)和工具变量模型相结合,利用 IV-Heckit 模型进行识别,以最大限度地消除回归分析中的内生性偏误。其中,农民工来源村庄的自然灾害强度被作为工具变量来识别收入模型。本文把从 OLS 模型,Heckit 模型和 IV-Heckit 模型分别得出的网络效应估计量概括在表 1 中。

表 1 OLS, Heckit 和 IV-Heckit 模型估算结果 (N = 2361)

	OLS	Heckit	IV-Heckit
同村打工网规模(ln)	. 125 *** (. 0349)	. 263 *** (. 076)	. 628 *** (. 232)
工具变量第一阶段回归 F 统计量	17. 42		
豪斯曼内生性检验	P = 0. 007		

注: (1) 括号内是标准误差。(2) * $p < 0. 1$, ** $p < 0. 05$, *** $p < 0. 01$ 。

表 1 中,第三列的估计值来自 Heckit 模型,也即考虑了可能的样本选择问题——外出打工本身就是一个选择性的过程。第四列的 IV-Heckit 模型是基于赫克曼二阶段法的工具变量模型,能够同时解决样本选择问题和一般的内生性问题。不难发现,IV-Heckit 模型给出了一个比 Heckit 模型更大的网络效应估计值:后者是 0. 263,而前者达 0. 628。如何解释 IV-Heckit 估计值大于 Heckit 估计值呢? 一个现成的

① 注意,在 LATE 框架下诠释工具变量估计值,需要的一个前提是自然灾害和网络之间的关系是单调的。

解释是,遗漏变量和网络大小正相关,但与个人收入负相关。但这样的解释是“空对空”的揣测,没有社会学意义。而如果考虑到自然灾害促使农民做出外出打工决定的“压力”不是均质的,那么就可以在 LATE 的框架下给出比较合理和直观的解释。因此,该文做了进一步的诠释:由于能力、地缘、历史习俗等因素,不同村庄的农民,在做出外出打工决定时,对自然灾害的敏感度可能是不一样的。例如,能力弱的村民或者平均能力较弱的村庄,对自然灾害造成的损失更加担忧,也就更容易被自然灾害“拉动”而外出打工。这样,“弱能力村庄”外出打工网的规模,就更容易受自然灾害影响。当我们用自然灾害作为工具变量来估算同村打工网的工资效应时,IV-Heckit 模型的估计值所体现的就不是基于样本的总体平均效应,而是一个加权平均值。其中,来自“弱能力村庄”的农民工会具有更大的权重。这一加权平均值,就是前面所谓的 LATE。而理论和实证研究都表明,能力弱的农民工,对基于村庄的同乡网依赖更多,从村庄网获得的边际效益也更大。也就是说,同村打工网的网络效应,在来自“弱能力村庄”的农民工群体中更强。既然 IV-Heckit 模型更多地反映了“弱能力村庄”中的网络效应,它给出的估计值自然就要比 Heckit 模型大。

五、工具变量分析实例: 社会网、选择性交友与求职

这一节我们通过采取虚拟实验来构建工具变量的实证分析案例,来展示一个标准的工具变量分析过程。通过实例展示,我们强调工具变量分析不仅仅能对因果推断做出贡献,更能够通过工具变量估计量和一般 OLS/Probit 估计量进行比较分析,使我们对社会机制及其过程产生深刻的理解和认识。解释和比较分析工具变量结果的过程,就是促使我们深入剖析社会机制的过程。

此项研究系国内社会学界一直比较关心的社会资本或社会网络课题,其直接目的在于检验一项经典假说:社会资本是否真有利于一个人找到工作(参见 Lin, 1999)。在研究中,社会资本的操作化定义是密友圈子的平均教育程度,因变量的操作化定义则是是否有工作。由于交友的过程往往是选择性的(经典的理论就是人以群分,参见 McPherson et al., 2001),因此朋友的平均教育程度可能是一个内生变量。这样,

一般的单方程模型就不能够证明社会资本的因果效用。

本研究数据来自 2003 年中国综合社会调查 (CGSS 2003)。该问卷中设计了被访者的核心讨论网以及被访者自身一系列人口学指标的相关调查。利用这些数据,我们进行工具变量分析,分“建立模型”、“寻找工具变量”、“数据分析”、“诠释比较”四个步骤。

1. 第一步: 建立模型

我们首先建立一个被访者是否找到工作的预测模型。考虑求职是二分变量,我们采用 Probit 模型,写方程如下:

$$P(y_i = 1) = \Phi(\beta_0 + \beta_1 S_i + \beta_2 X_i) \quad (4)$$

其中 $y_i = 1$ 表示被访者 i 找到工作, S_i 表示社会资本,也即被访者 i 的朋友的平均教育程度, X_i 是一系列的外生个人特征变量,如性别、年龄等。

方程 (4) 实际可以写作另外的形式:

$$\hat{S}_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_i + \hat{\gamma}_2 X_i \quad y_i = 1 \text{ if } y_i^* > 0 \quad (5)$$

其中 y_i^* 是一个潜在变量,也即如果 $y_i^* > 0$ 则 $y_i = 1$ 。而工具变量模型 (IV-Probit) 可以用下面的方程组来表示:

$$y_i^* = \beta_0 + \beta_1 S_i + \beta_2 X_i + \varepsilon_i, y_i = 1 [y_i^* > 0] \quad (6)$$

$$S_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 X_i + \xi \quad (7)$$

这里, Z_i 就是工具变量, ξ 是随机误差项。这里, $\text{Cov}(Z_i, \varepsilon_i) = 0$, $\text{Cov}(\xi, \varepsilon_i) = 0$, 且 $\text{Cov}(Z_i, S_i) \neq 0$ 。

步骤要点: 必须清楚地说明对模型进行无偏估计所需要的全部假设。同时,第一阶段回归方程里(也即方程 7),要把第二阶段方程(也即主方程 6)的全部外生变量都要放入,哪怕直觉上这些外生变量和方程左侧的内生变量没有关系。

2. 第二步: 寻找工具变量

一般而言,我们都是从调查数据中寻找可能的工具变量。这是因为,凭空想象出一个有效的工具变量是非常难的。而运用逆向思维,从数据能提供的一系列变量里进行逐一排查和联想,则可能寻找到可用的工具变量。不过,本研究采用的是一种基于模型工作假设进行虚拟实验以人工构建工具变量的新方法。该方法在计量经济学文献中已经得到运用,其主要目的就是用来消除选择性偏误(如 Bayer & Ross, 2009)。当然,如果我们采取一般的寻找工具变量的方法,那么本部分内容读者可以略去并直接跳转到第三步。

要对方程(5)进行无偏估计,我们的标准假设是 S_i 和 X_i 都是外生的。但由于交友的过程很可能是选择性的,因此我们将假设放松为 S_i 是内生的,而 X_i 仍然是外生的,也即 $\text{Cov}(X_i, \varepsilon_i) = 0$ 且 $\text{Cov}(S_i, \varepsilon_i) \neq 0$ 。也就是说,朋友圈的教育程度是内生的,其他控制变量是外生的。利用这一工作假设,我们可以巧妙地构造出一个虚拟的工具变量。

首先,我们知道 X_i 代表的这些个人特质基本都具有离散型的特点,例如被访者的个人教育程度、性别、年龄段等等。为构建工具变量,我们为每个被访者 i 分配一批“虚拟朋友”。匹配的原则是,这些“虚拟朋友”都是除 i 之外的其他被访者,但同时具有和 i 相同的个人特质。这样,在每个虚拟的朋友圈子里,他们都具有相同的 X ,然后,对每一个被访者 i 我们对其“虚拟朋友”所各自拥有的现实朋友圈的教育程度取平均值。而这个平均值,就是我们的工具变量。其合法性体现在如下两个方面:(1)被访者 i 的现实朋友圈教育程度(S_i)和他的“虚拟朋友”的现实朋友圈教育程度平均值(Z_i)相关。这是因为被访者 i 和他的“虚拟朋友”具有相同的个人特性。而根据社会同质性理论(social homophily)相似的人往往会在交友时作出相似的选择。这一相关性及其强度,可以在2SLS的第一阶段回归中得到统计验证。(2)被访者 i 的“虚拟朋友”的现实朋友圈教育程度平均值(Z_i)必然和误差项 ε_i 无关。这是因为,该工具变量的形成过程,没有基于任何非观测因素或者自选择。也就是说,“虚拟朋友圈”的形成,根据且仅仅根据的是一组既定的外生控制变量 X_i 。而 X_i 的外生性,是我们的标准工作假设所规定了的。因此,只要我们从一开始就假定 X_i 是外生的,则该工具变量也必然是外生的。实际上,我们可以把这个工具变量理解成一个用非参数的选择模型来预测出的每个被访者 i 的朋友圈教育程度。

那么,如何理解我们构建工具变量时所依赖的工作假设 $\text{Cov}(X_i, \varepsilon_i) = 0$,^①也即控制变量是外生的假设?在假设的基础上人为构建一个工具变量,是不是使得工具变量的外生性不具备足够的说服力?在这一点上,正如巴耶尔和罗斯(Bayer & Ross, 2009)所强调的“如果我们的分析仅希望把因为自选择而导致的偏误消除掉,而不是野心勃勃地宣称要消除全部的内生偏误,那么在标准工作假设的基础上构造工

① 例如,有人会认为,诸如个人教育程度和政治面貌等控制变量也会是内生的。不过,控制变量被假定为外生,是我们用回归模型进行任何一种因果分析的标准假设。

具变量是有充分理由的”。实际上,本研究给出的工具变量估计量(IV-Probit 估计量)之所以比单方程的 Probit 估计量更具有意义,是因为它消除了基于干扰变量的自选择问题,而不是说它是无偏的。

在本研究中,我们使用以下几个离散的控制变量来构建“虚拟朋友圈”:年龄段(16-30, 30-40, 40-60);性别(男,女);教育程度(小于13年,13-16年,17年以上);父亲政治面貌(党员和非党员);父亲的行政管理级别(无级别;市、区、县;省级;中央部委),以及经济地理区划(根据GDP划分为四个等级)。显然,这6个控制变量可以产生 $3 * 2 * 3 * 2 * 4 * 4 = 576$ 个“虚拟朋友圈子”。全部被访者都被分配到这576个虚拟圈子之中。圈子平均大小是33。在这个基础上,我们对每个被访者的“虚拟朋友圈子”的全部成员的朋友圈教育程度进行取平均值,以此得到工具变量(此处略去数据描述)。

步骤要点:因为工具变量的外生性是无法直接进行统计证明的,所以在这一步中我们应该尽可能全面、详细地向读者论证其外生性,要做到令人信服;关于工具变量和解释变量的相关性,则可以不过多着墨。因为,我们可以在下一步骤中用第一阶段F统计量来证明。

3. 第三步: 数据分析

有了工具变量,我们就可以进行基于2SLS的回归分析。单方程Probit模型估计量和IV-Probit模型估计量见表2。

在分析IV-Probit估计量之前,我们先分析第一阶段回归的F统计量。其值为38.3,显然大于经验值10,故此不存在弱工具变量问题。也就是说,我们通过虚拟实验构建的工具变量和解释变量是紧密相关的(基本特征类似的被访者,在交友上具有一定的趋同性)。而瓦尔德内生性检验表明IV-Probit模型和Probit模型存在系统差异,因此要采用一致的工具变量估计量。因篇幅原因,对表2中各系数的基本分析、各类稳健性测试和第一阶段回归结果文中略去。

步骤要点:必须非常清楚地告诉读者有关关键统计量,以证实进行工具变量分析的必要性和工具变量的质量。最好能在表格中同时把单方程OLS/Probit模型回归的结果也纳入,以便在下一步和多方工具变量模型的结果进行比较分析。

4. 第四步: 机制诠释

从表中不难发现,IV-Probit估计量不但统计上显著,其大小也几乎是Probit估计量的三倍多。这样,基于有关工作假设,本研究得出的技

表 2 求职模型估计量

自变量	Probit 单方程模型	IV-Probit 模型
朋友圈的平均教育年份	.066 ***	.26 ***
年龄	.044 ***	.060 ***
年龄的平方	-.00068 ***	-.0009 ***
教育年份	.086 ***	.026 * *
父亲是党员	.18 ***	.068
父亲单位是市区县级	.188 ***	.202 ***
父亲单位是省级	.203 ***	.217 ***
父亲单位是中央级	.076	.124
女性	-.589 ***	-.54 ***
经济地理区划	YES ***	YES ***
截距	-1.79	-2.88
第一阶段 F 统计量	-	38.3
瓦尔德内生性检验	-	Chi2(1) = 5.69 P = 0.017
样本数	3974	3974

注: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

术性结论就是双重的: 第一, 朋友圈的教育程度对求职有作用; 第二, 原先的 Probit 模型低估了朋友圈教育程度对求职的作用。那么, 如何解释 IV-Probit 估计量和 Probit 估计量之间的差异? 我们的诠释是: 工具变量估计量之所以更大, 很可能是因为“差异性交友”的存在 (heterophilous selection) 而这与我们所常见的人以群分的“趋同性交友” (homophilous selection) 完全相反。下面做一详细诠释。

如果交友的选择性是出于“同质性” (homophily), 也即“人以群分”, 那么普通 Probit 模型应该会高估——因为有偏的估计量实际上代表了社会资本和“能力”、“口才”、“性格”这一类干扰项的联合作用。反之, 如果人们因为某种社会条件限定而采取特殊的“补偿策略” (compensation strategy), 比如就业前景差的人主动和前景好、学历高的人交往, 那么普通 Probit 估计量就从真实的网络效应中抵消了一些干扰项的作用, 也就会比 IV-Probit 估计量小。因此, 在这项研究中, 我们的实质性诠释就是: 就业能力或前景差的人, 可能的确采取了一种补偿策略, 在择友过程中努力结识文凭较高的人以获取社会资源。

作为一种诠释,“差异性交友”虽然可以解惑,但我们难免要就这一诠释背后的社会机制进一步追问。比如,“差异性交友”何以可能?毕竟我们长期以来对“人以群分”耳熟能详,而后者,也更符合我们的直觉。因此,我们必须从纯技术性的解读和揣测转向寻求更深层的社会学理论支持,以证明我们对工具变量估计量的诠释具有合理性。

虽然“趋同性交友”是众所周知的现象,但在实际的社会网络形成过程中,其程度和范围都未必有我们想象得那么深广。麦克费森等人(McPherson et al. 2001)的经典研究就承认,尽管在种族、性别、宗教、年龄方面“趋同性交友”很明显,但在教育、职业、社会网位置等方面则未必如此。特别是对于密友关系,教育程度的相似性实际上并不重要(Louch 2000)。因此,正如林南(Lin 2008)所说,同质或异质网络并不是对立的选择,选择朋友的动机非常复杂。从理论角度,我们起码可以为“差异性交友”的可能性提供四个方面的依据。

第一,“差异性交友”可能是工具性行动的表现形式。由于社会资源具有不平等性,在情感表达等社会关系中,“趋同性交友”比较普遍;但在寻求社会资源的工具性行动中,“差异性交友”占优势,并且能改善社会资源分布不均的情况(Lin 2008; Briggs, 1998, 2002; Wellman & Gulia, 1999; Dominguez & Watkins, 2003)。本研究的数据来自CGSS 2003问卷,而我们从相关数据得知,作为密友圈子的核心讨论网,其所讨论的内容,有70%是和求职有关的。因此,这个核心讨论网,具有明显的工具性行动的特征。

第二,“差异性交友”形成的不对称关系可以是维系的。这种不平等的关系之所以可以维系下去,其原因就在于:“低位”者有可能愿意采取服从、跟随的姿态以消除内心情绪并获取资源,而“高位”者虽不能从关系中获得社会资源,但可以获得特定的满足感和成就感(Robins & Boldero 2003; Dominguez & Watkins 2003)。

第三,“差异性交友”往往由特定的社会背景和交往机会决定。例如,布劳(Blau, 1977)就指出,个人社会关系不仅仅是个人的愿望,更取决于社会经济结构所决定的交往机会。一个金字塔型社会中,高位者总是少数,社会资源的过度贫乏也会导致人们努力向上追求建立网络。

第四,“差异性交友”之所以出现在教育指标上,很可能是因为教育程度本身在转型期的中国,仍然不是社会资源最直接的表征。例如,“趋同性交友”可能会出现在权力、财富等直接资源的指标上:一个本

科毕业的科长、处长,比一名拥有博士学位的教授、科学家的权力更大。也就是说,“差异性交友”很可能只是“弱者的选择”,也是无可奈何的选择。

总结这一工具变量分析的实例,我们就不难理解摩根(Morgan, 2002)的一个重要观点,那就是:如果把工具变量分析仅仅作为因果推断的工具,未免太狭隘了。而通过对工具变量估计量和IV-Probit估计量的比较分析,我们就可以对选择性友谊的形成有更深入的了解。这虽然与社会资本因果判断没有直接关系,但对于深化研究极为重要。

步骤要点:在诠释结果时应该尽量对工具变量估计量和OLS或Probit估计量进行比较分析,以探寻偏误的来源和方向,并寻找合理的社会学理论和实际社会现象来解释这些差异。没有社会学实质意义的工具变量诠释,会使得我们的工具变量研究沦为纯粹的数学技巧展示。

六、结语:局限抑或潜力?

既然工具变量方法功能如此强大、简洁明了,那么,为何社会学家还是一直敬而远之?笔者认为,这既因为学科方法的演进本身具有路径依赖的特征,同时也因为我们对工具变量局限性的认识仍然不够全面。这里,笔者针对加诸工具变量的质疑,^①逐一进行客观分析回应。

第一,找到好的工具变量非常之难,但难寻不该是不去尝试的理由。工具变量难寻就意味着该方法对数据本身有很大的依赖。如果数据不能提供合适的工具变量,我们就面临无米之炊的局面。但是,工具变量难寻,不应成为社会学家不去探索工具变量的理由。它难寻的原因,恰恰就是它的力量所在:因为它必须永远置身模型之“外”。当我们一心关注模型本身之时,我们的注意力自然只会放在模型之“内”。因此,只要我们变换思维的角度,开始习惯从模型之“外”寻找解决问题的武器,多从前人巧妙使用工具变量的实例里获得启发,我们就会发现,工具变量虽然难寻,但绝非了无痕迹。实际上,摩根(Morgan,

① 较新的综合性文献回溯可以参见迪顿对计量经济学特别是发展经济学中基于自然实验方法的工具变量运用的批评(Deaton, 2010)。除了对外生性的怀疑、局部干预效应的不利于诠释等方面,迪顿主要认为实验和自然实验工具变量方法往往难以具有我们期望的政策意义。

2002) 对社会学家提出了一个非常中肯的意见: 只要和其他方法估计量进行比较和相互补充, 就可以更大胆、更冒险地去发现和使用工具变量, 哪怕其外生性有明显瑕疵或无法完全消除质疑, 也是值得的。为此, 他专门以自己用倾向值匹配方法进行教育回报研究的例子, 说明了即使是有问题的工具变量, 也会对倾向值匹配方法估计带来有益和重要的补充。在这个意义上, 问题不在于有没有工具变量, 而在于要善于把工具变量和其他模型相互结合、比较。

第二, 工具变量的外生性无法直接用统计方法来验证, 但实际上几乎所有的定量分析方法都基于诸多强假设。工具变量对于模型的外生性实际是无法用统计方法检验的。这一尴尬直接导致的结果就是: 不管熟悉还是不熟悉工具变量方法的读者, 面对一篇工具变量分析的论文, 其第一反应总是首先质疑其合法性。一旦这一关通不过, 整个后续的分析在读者眼中就失去了说服力。但凭心而论, 和其他一些解决内生性问题的常用模型相比, 工具变量分析承受了过多的挑刺和压力。当固定效应模型假设干扰项不随时间变化, 当倾向值匹配模型假设一切偏误都来自可观察的变量, 读者们往往能不假思索地接受这些武断的假设。原因何在? 这是因为, 无论是固定效应模型还是倾向值匹配模型, 确保它们可以运行的工作假设, 往往在论文中被预设于数据分析之前, 而对假设一旦不能满足所带来的问题和解释力的局限性, 则在文章最后一笔轻轻带过甚至不加提及。但工具变量的外生性作为一个工作假设, 哪怕已经有充分的直觉、理论和实证支持, 但却因为直接进入了分析过程之中而显得格外扎眼。有时候, 学者宁可近乎偏执地争论地震是不是真正地“外生”(如 Deaton, 2010), 而对固定效应和倾向值匹配模型中假设个人异质性恒定不变、偏误来自可观测的干扰项等“强假设”不置一词。

第三, 工具变量估计量往往因工具变量的选取而异, 但细化的“局部效应”分析往往能够为我们带来对社会机制的深刻认识。工具变量对样本的影响, 往往是非均质的, 所以工具变量估计量带有权重性的特征, 即前文提及的局部干预效应。这样, 从工具变量分析里得到的结论, 往往“适用”于样本中的一部分, 有时候甚至很难诠释, 因此降低了社会科学分析的政策意义(Deaton, 2010)。不过, 摩根(Morgan, 2002)独辟蹊径地指出, 与其说这是工具变量方法的不足, 倒不如说是一种优势。因为基于LATE框架下的工具变量分析, 样本变得更加具有目标

性结论也更有说服力。从本文所举的农民工收入的网络效应例子来看,起码我们可以有把握地说:同村外出打工网,在能力较弱的一部分人中,具有很强的收入效应。当然,如果在LATE框架之下解释过度识别的模型,解释的难度确实会增加很多。特别是,为满足工具变量和解释变量之间关系单调的条件,我们不应在模型中直接联合使用工具变量,而应该各自使用,各自诠释。

第四,工具变量的使用看似难以举一反三,但哪怕是前人使用过的失败的工具变量,都可能具有借鉴和推广意义。美国的教育回报研究中将地理距离作为工具变量的方法,看起来并不适用于中国。在中国的教育回报和社会网研究中,文化大革命、上山下乡等政策背景可以作为工具变量,这也无法推广到其他国家的不研究中去。但是,并非所有的工具变量都无法推广。例如,降雨量、自然灾害等作为外生工具变量,无论在哪个国家哪个地区,都会对社会进程产生外生的影响。甚至被诟病过的弱工具变量,在不同的研究情境中也有可能成为一个好的工具变量。前文提及曾有美国研究者把出生季度作为教育年份的工具变量(Angrist & Krueger, 1991),但后来研究者发现,出生季度的不同所导致的教育年份差异是非常微弱的,因此这是个弱工具变量(Bound et al., 1995)。不过,吴要武(2010)认为,美国数据中被访者的高中毕业比例一般非常高,因此出生季度导致的1年之差所占个人总教育年份的比例非常小。而在不发达国家由于高中毕业比例非常低,情况就会大有不同。因此,他利用2005年中国人口1%抽样调查数据,有力地证明了在中国的教育回报研究中,出生季节是受教育年份的一个非常好的工具变量,其F统计量达到30多。

尽管如此,在结束本文之前,仍然有必要提醒:工具变量分析者必须始终保持审慎的头脑。摩根在与笔者的讨论中,多次提及工具变量分析论文常有的过份自信(oversell)的情形。其主要原因,一是研究者本人确信了工具变量的外生性之后,难免会有一种自负,认为可以充分解决内生性问题,从而忽视很多细节;二是工具变量的寻找确实是一种高度的智力挑战,自以为找到圣杯的人自然会欢欣鼓舞。但是,一个人的思维力量往往是有限的。计量经济学中那些大名鼎鼎的经典案例,不管是使用抽签服役、河流数量、殖民地死亡率还是出生季度作为工具变量,在研究发表多年之后仍然受到不断的挑战。发现和使用工具变量,既能展示研究者具有一定的社会科学逻辑力、想象力和诠释力,但

同时也会把研究最薄弱的环节(无法验证的外生性)直接展示给学术界。因此,工具变量既是展示社会科学想象力的舞台,也是可以让一篇论文失去价值的达摩克利斯之剑。从这个角度上来说,所有的工具变量方法使用者,都应该小心谨慎,清楚地说明所需假设,清楚地说明一旦工具变量外生性假设不能满足,估计量会发生偏移的方向。只有这样,才能把理论上的直觉和前人使用工具变量的经验教训结合起来,让我们的逻辑、想象和诠释,为社会科学定量分析中的因果推断提供更强的说服力和更高的可信度。

参考文献:

- 陈云松 2012,《农民工收入与村庄网络:基于多重模型识别策略的因果效应分析》,《社会》第4期。
- 陈云松、范晓光 2010,《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》第4期。
- 2011,《社会资本的劳动力市场效应估算:关于内生性问题的文献回溯和研究策略》,《社会学研究》第1期。
- 方颖、赵扬 2011,《寻找制度的工具变量:估计产权保护对中国经济增长的贡献》,《经济研究》第5期。
- 胡安宁 2012,《倾向值匹配与因果推论:方法论述评》,《社会学研究》第1期。
- 梁玉成 2010,《社会资本和社会网无用吗》,《社会学研究》第5期。
- 魏万青 2012,《户籍制度改革对流动人口收入的影响研究》,《社会学研究》第1期。
- 吴要武 2010,《寻找阿基米德的“杠杆”:“出生季度”是个弱工具变量吗》,《经济学(季刊)》第2期。
- 谢宇 2006,《社会学方法与定量研究》北京:社会科学文献出版社。
- 章元、陆铭 2009,《社会网络是否有助于提高农民工的工资水平》,《管理世界》第3期。
- Acemoglu, D., S. Johnson & J. A. Robinson 2001, “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review* 91.
- Angrist, J. D. 1990, “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.” *American Economic Review* 80.
- 1993, “The Effects of Veterans Benefits on Education and Earnings.” *Industrial and Labor Relations Review* 46.
- Angrist, J. D. & W. N. Evans 1998, “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” *American Economic Review* 88(3).
- Angrist, J. D. & A. B. Krueger 1991, “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics* 106(4).
- 1992, “Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery.” NBER Working Paper 4067.

- 2001, “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments.” *The Journal of Economic Perspectives* 15(4) .
- Angrist, J. D. & V. Lavy 1999, “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114(2) .
- Angrist, J. D. , G. W. Imbens & D. B. Rubin 1996, “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91.
- Angrist, J. D. , S. H. Chen & B. R. Frandsen 2010, “Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-reported Health.” *Journal of Public Economics* 94(11-12) .
- Bayer, P. & S. L. Ross 2009, “Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application.” NBER Working Paper 12211.
- Bentolila, S. , C. Michelacci & J. Suarez 2010, “Social Contacts and Occupational Choice.” *Economica* 77(305) .
- Blau, P. M. 1977, *Inequality and Heterogeneity*. New York: Free Press.
- Bobonis, G. J. & F. Finan 2008, “Neighborhood Peer Effects in Secondary School Enrollment Decisions.” Working Paper.
- Booser, M. & S. E. Cacciola 2001, “Inside the ‘Black Box’ of Project Star: Estimation of Peer Effects Using Experimental.” Working Paper.
- Bollen, K. 2012, “Instrumental Variables in Sociology and the Social Sciences.” *Annual Review of Sociology* 38(22) .
- Bound, J. , D. A. Jaeger & R. Baker 1995, “Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak.” *Journal of the American Statistical Association* 90(430) .
- Bramoullé, Y. , H. Djebbari & B. Fortin 2007, “Identification of Peer Effects Through Social Networks.” IZA Discussion Paper 2652.
- Briggs, Xavier de Souza 1998, “Brown Kids in White Suburbs: Housing Mobility and the Multiple Faces of Social Capital.” *Housing Policy Debate* 9(1) .
- 2002, “Social Capital and Segregation: Race, Connections, and Inequality in America.” Working Paper RWP 02-011, Kennedy School of Government, Harvard University.
- Card, David 1995, “Using Geographic Variation in College Proximity to Estimate the Return to Schooling.” In L. N. Christofides, E. K. Grant & R. Swidinsky (eds.) , *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.
- 1999, “The Causal Effect of Education on Earnings.” In O. Ashenfelter & D. Card (eds.) , *Handbook of Labor Economics* (Vol. 3A) . Amsterdam: North Holland.
- Card, D. & A. 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(4) .
- Cipollone, Piero & Alfonso Rosolia 2007, “Social Interactions in High School: Lessons from an

- Earthquake. ” *American Economic Review* 97(3) .
- Contreras ,D. ,D. Kruger ,M. Ochoa & D. Zapata 2007, “The Role of Social Networks in the Economic Opportunities of Bolivian Women. ” Working Paper.
- Cutler , David & Edward Glaeser 1997, “ Are Ghettos Good or Bad?” *Quarterly Journal of Economics* CXII.
- Deaton , Angus 2010, “Instruments , Randomization , and Learning about Development. ” *Journal of Economic Literature* 48(2) .
- Domínguez , S. & C. Watkins 2003, “Creating Networks for Survival and Mobility: Social Capital Among African-American and Latin-American Low-income Mothers. ” *Social Problems* 50.
- Durlauf , S. N. 2002, “On the Empirics of Social Capital. ” *Economic Journal* 112(483) .
- Durlauf , S. N. & M. Fafchamps 2004, “Social Capital. ” NBER Working Paper 10485.
- Evans , W. N. , W. E. Oates & R. M. Schwab 1992, “Measuring Peer Effects: A Study of Teenage Behavior. ” *Journal of Political Economics* 100.
- Grogger , Jeff 1996, “School Expenditures and Post-schooling Earnings: Evidence from the High School and Beyond. ” *Review of Economics and Statistics* 78(4) .
- Hall , R. E. & C. I. Jones 1999, “Why Do Some Countries Produce So Much More Output per Worker than Others?” *Quarterly Journal of Economics* 114(1) .
- Hanushek , E. A. , S. G. Rivkin & L. L. Taylor 1996, “Aggregation and The Estimated Effects of School Resources. ” *Review of Economics and Statistics* 78(4) .
- Hoxby , Caroline 2000, “Does Competition Among Public Schools Benefit Students and Taxpayers?” *American Economic Review* 90(5) .
- Imbens , Guido W. & J. D. Angrist 1994, “Identification and Estimation of Local Average Treatment Effects. ” *Econometrica* 62(2) .
- Lin , N. 1999, “Social Networks and Status Attainment. ” *Annual Review of Sociology* 25.
- 2008, “A Network Theory of Social Capital. ” In D. Castiglione , Jan van Deth & G. Wolleb (eds.) , *The Handbook of Social Capital*. Oxford and New York: Oxford University Press.
- Louch , Hugh 2000, “Personal Network Integration: Transitivity and Homophily in Strong-tie Relations. ” *Social Networks* 22.
- Manski , Charles F. 1993, “Identification of Endogenous Social Effects: The Reflection Problem. ” *Review of Economic Studies* 60(3) .
- 2000, “Economic Analysis of Social Interactions. ” *Journal of Economic Perspectives* 14.
- Maurin E. & J. Moschion 2009, “The Social Multiplier and Labor Market Participation of Mothers. ” *American Economic Journal: Applied Economics* 1(1) .
- McPherson , M. , Lynn Smith-Lovin & J. Cook 2001, “Birds of a Feather: Homophily in Social Networks. ” *Annual Review of Sociology* 27.
- Moffitt , R. A. 1995, “Selection Bias Adjustment in Treatment-effects Models as a Method of Aggregation. ” In 1996 Proceedings of the American Statistical Association.
- 2001, “Policy Interventions , Low-Level Equilibria and Social Interactions. ” In S. N. Durlauf & P. H. Young (eds.) , *Social Dynamic*. Boston , MA: MIT Press.

- Morgan , S. L. 2002, “Should Sociologists Use Instrumental Variables?” Working Paper , Cornell University.
- Morgan , S. L. & C. Winship 2007 , *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Munshi , Kaivan 2003, “Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. ” *Quarterly Journal of Economics* 118(2) .
- Mouw , Ted 2003, “Social Capital and Finding a Job: Do Contacts Matter?” *American Sociological Review* 68(6) .
- 2006, “Estimating the Causal Effect of Social Capital: A Review of Recent Research. ” *Annual Review of Sociology* 32.
- Qian , Nancy 2008, “Missing Women and the Price of Tea in China: The Effect of Sex-specific Earnings on Sex Imbalance. ” *Quarterly Journal of Economics* 123(3) .
- Rivkin , Steven 2001, “Tiebout Sorting , Aggregation and the Estimation of Peer Group Effects. ” *Economics of Education Review* 20(3) .
- Robins , G. & J. Boldero 2003, “Relational Discrepancy Theory: The Implications of Self-discrepancy Theory for Dyadic Relationships and for the Emergence of Social Structure. ” *Personality and Social Psychology Review* 7(1) .
- Sovey , A. J. & D. P. Green 2011, “Instrumental Variables Estimation in Political Science: A Readers’ Guide. ” *American Journal of Political Science* 55(1) .
- Stock , J. H. & F. Trebbi 2003, “Retrospectives: Who Invented Instrumental Variable Regression?” *Journal of Economic Perspectives* 17(3) .
- Stock , J. H. & M. Yogo 2005, “Testing for Weak Instruments in Linear IV Regression. ” In J. H. Stock & D. W. K. Andrews (eds.) , *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press.
- Wellman , Barry & Milena Gulia 1999, “The Network Basis of Social Support: A Network Is More Than the Sum of Its Ties. ” In Barry Wellman (ed.) , *Networks in the Global Village: Life in Contemporary Communities*. Boulder , CO: Westview.

作者单位:南京市玄武区政府
责任编辑:杨典

corresponding actions taken during the campaign were based on their own benefits , values , and emotions. Thus , peasants’ voice in *suku* was not always the reproduction of the narrative mode of class suffering , but multiple choices including acceptance , imposition , alteration , avoidance , reticence and even resistance. This paper emphasizes the selective adaption of peasants as leading participants to the limited freedom of speech during *suku*. The authors argue that with this selective adaption , *suku* was adopted by peasants to maintain their own interests and emotional needs , while at the same time , strategic answers were given to cope with the prohibition of voice violating the narrative mode of class suffering. Therefore , CCP’ s political discipline did show great effect in short term , but inevitably failed at last. Furthermore , on the contrary to expectation , the utility of *suku* also started to diminish over time.

Subsistence Ethic and Collectivity Logic: A sociological analysis of “*Daoqianhu*” phenomenon in the period of agricultural collectivization
 *Meng Qingyan* 172

Abstract “*Daoqianhu*” , as a unique historical phenomenon , emerged in the middle of 1950s when the agricultural collectivization movement got started. The existence of “*daoqianhu*” illustrated the state’s effort to reconcile the conflict between the peasants’ subsistence ethic and the maximization of individual incentives. Meanwhile , it was also a power practice that the CCP tried to create the new socialist generation. However , the state power didn’t achieve its institutional goal of rebuilding the rural society. Based on the review of existing theoretical and empirical researches , this paper emphasizes that we should understand “*daoqianhu*” from ethical point of view rather than the overly simplified economic angle. The author argues that the production team , which was seen by peasants as a family-like organization , was not only an economic organization but also a social organization , which resulted into a collectivity ethic that confounds the public and the private. Together with other factors , it led to the failure of “*daoqianhu*” institution and facilitated other unexpected historical results.

Logic , Imagination and Interpretation: The application of instrumental variables for causal inference in the social sciences
 *Chen Yunsong* 192

Abstract: Instrumental variable (IV) method is one of the most important strategies dealing with the endogeneity problem in quantitative analysis to conduct causal inference. This paper introduces the concept , principle and model of IV , and reviews five types of IVs used in previous studies. It also illustrates the Local Average Treatment Effect which is of importance to interpret the IV estimates. An empirical

analysis is employed as a template to demonstrate how to conduct an IV analysis. Finally , the criticism of IV method is also discussed.

REVIEW

Virtual Ethnography: The field , approach and ethics
..... *Bu Yumei* 217

Abstract: The internet and its derived social and cultural phenomena not only have brought a new research topic to anthropology ,but also have challenged the traditional anthropological research methods. Based on the adaption and improvement of traditional ethnographic approach to fit the new research setting , virtual ethnography was created to comprehend the internet and its related phenomena. Starting from the articulation of the association between the internet and the culture ,the author reviews and discusses several issues important to virtual ethnography , such as the definition of virtual field , the realization of online participant observation , the expansion of online to offline fieldwork , and research ethics. Finally , the author argues that reflecting the tradition as well as maintaining the dialogue with tradition is an important approach for the further development of virtual ethnography.