

倾向值匹配与因果推论： 方法论述评^{*}

胡安宁

提要: 社会科学对于因果关系的关注使得倾向值匹配(propensity score matching)的方法受到越来越多的重视。本文从以下几个方面对这一方法进行述评。首先,本文系统梳理了倾向值匹配方法的历史、发展及其对于社会学调查研究的意义。其次,通过一个具体实例,本文展示了倾向值匹配的具体实施过程。第三,从密尔对因果关系的论述以及统计学的反事实框架(counterfactual framework)出发,本文阐释了倾向值匹配如何通过“控制”混淆变量从而满足非混淆假设(unconfoundedness)来帮助研究者提出因果性结论。第四,本文将倾向值匹配与回归中断设计和工具变量进行了比较。最后,在简要分析倾向值匹配的局限性之后,本文讨论了倾向值匹配的最新发展趋势。

关键词: 因果推论 倾向值匹配 控制思想 反事实框架

在过去十几年中,社会科学领域中一个重要的方法论突破在于开始关注如何通过严格的统计技术进行因果推论(Sobel,1995,1996,2000;Morgan & Winship,2007;Winship & Morgan,1999;Winship & Sobel,2004)。在各种统计学方法中,倾向值匹配(propensity score matching)受到越来越多的重视,并在教育学、传染病学以及社会学等领域得到了广泛的应用(Harding,2003;Morgan,2001;Morgan & Harding,2006;Normand et al.,2001;Smith,1997)。

相对于经济学的工具变量方法,倾向值匹配在社会科学研究中还是一个新兴的统计技术。然而,随着越来越多的经验研究开始关注并且应用倾向值匹配,其独特的优势也逐渐展现出来,因此有着极为广阔的应用前景。本文将从六个方面系统介绍这一方法。在第一部分,笔者将讨论倾向值匹配的历史、发展以及它对社会学调查研究的意义。

^{*} 作者感谢复旦大学社会学系周怡教授和北京大学社会学系方文教授的建议和帮助,并感谢匿名审稿人所提出的修改意见。文责自负。

第二部分将通过一个具体实例来展示倾向值匹配的具体实施过程。在第三部分,笔者将从科技哲学和统计学两个方面讨论为什么倾向值匹配的结果可以进行因果性解读。第四部分将把倾向值匹配与其他比较常用的因果推论方法进行简要对比并由此展示倾向值匹配相对于其他方法所特有的优势。在第五部分,笔者将探讨倾向值匹配所存在的一些局限性。最后,简要总结本文内容并探讨倾向值匹配的最新发展趋势。

一、倾向值匹配: 历史、发展及其对调查研究的意义

倾向值(propensity score)这一概念最早出现在1983年罗森鲍姆和鲁宾合写的一篇名为《倾向值对于观察研究中因果效应的中心作用》的论文中(Rosenbaum & Rubin, 1983)。倾向值指被研究的个体在控制可观测到的混淆变量(confounding variables)的情况下受到某种自变量影响的条件概率。在一般的实证研究中,由于存在很多其他变量混淆自变量和因变量之间的关系,研究者很难直接探索二者之间的“净效果”(net effects)。例如,大学教育的收入回报可能受到年龄、性别以及地区差异这些混淆变量的影响。因此,上大学的人的平均收入与没有上大学的人的平均收入的差异有可能也反映了混淆变量的作用。这些混淆变量的影响通常被称为选择性误差(selection bias),而通过倾向值匹配的方式来控制和消除选择性误差则是罗森鲍姆和鲁宾论文的题中之义。罗森鲍姆和鲁宾用数理方法证明:通过将这些混淆变量纳入Logistic回归模型来产生一个预测个体受到自变量影响的概率(倾向值),研究者可以通过控制倾向值来遏制选择性误差对研究结论的影响从而保证因果结论的可靠性。

在各种控制倾向值的方法中,匹配(matching)比较简便易行。其基本逻辑是将受到自变量影响的个体与没有受到影响的个体进行配对。而倾向值匹配就是保证匹配起来的个体的倾向值相等或者近似。再次回到大学教育的例子,倾向值匹配就是将没有受过大学教育的人和受过大学教育的人配对并保证他们的倾向值(即上大学的概率)相同或近似,这样我们就得到了上过大学和没上过大学两组个体。因为已有的混淆变量已经在基于倾向值的配对过程中被控制起来了,两组

个体收入上的差异就只能归因于大学教育的有无,而不是其他混淆变量,由此遏制了选择性误差。

罗森鲍姆和鲁宾的思想与计量经济学家赫克曼的想法不谋而合。从样本选择的角度出发,赫克曼认为社会科学家在研究某个自变量效果的时候会不自觉地选择一些样本而忽视另外一些。例如,针对某个研究项目,不同的个体有不同的参与意愿,或者研究者对于研究个体的选择有特定的倾向。这些选择效应(selection effect)使得研究者真正接触的样本是有一定“偏向的”。而基于一个带有选择性的样本,我们得出的结论也会由于选择效应的“污染”而出现偏差(Heckman, 1979; Heckman & Smith, 1995)。面对这种情况,赫克曼认为在进行研究时需要做两步工作。第一步是建立模型来拟合个体参与到某种研究项目中的可能性,第二步则是在控制这种参与可能性的前提下考察我们所关心的自变量对因变量的效果。因此,赫克曼的方法也被称为“两步估计法”(two-step estimator)(Heckman, 1978, 1979)。

我们今天所探讨的倾向值匹配可以说是罗森鲍姆和鲁宾的统计学传统与赫克曼的计量经济学传统的结合。罗森鲍姆和鲁宾通过逻辑回归方法将混淆变量“总结”成一个倾向值,而这也恰恰是赫克曼两步估计法中的第一步所要做的,即通过模型估算个体参与某项研究的概率,二者殊途同归。然而,这两个传统的一个不同点在于他们处理倾向值的方式有一定差异。在得到倾向值之后,罗森鲍姆和鲁宾提出了多种控制倾向值的方法:除了匹配之外,也可以将倾向值分层(stratification)或细分(sub-classification),然后在每个层次或类别内探索因果关系;或者也可以将倾向值作为新的变量进行回归分析。^①而在赫克曼的计量经济学传统中,倾向值主要是作为一个新的变量被纳入到已有的计量经济模型中,例如选择模型(selection model),工具变量(instrumental variable),或双重差分方法(difference-in-differences method)。需要指出的是,这种差异只是在处理倾向值的具体方式上的不同,本质上这两个传统的基本逻辑是一致的,即在考察自变量效果之前通过模型估计并控制个体受到自变量影响的概率。

在社会学领域,早在1997年,宾夕法尼亚大学社会学系的史密斯就已经使用倾向值匹配方法研究医院改革对患者死亡率的影响

^① 将倾向值纳入回归模型的方法被称为“回归调整”(regression adjustment)。

(Smith, 1997)。然而,史密斯这篇文章的主要目的是展示倾向值匹配方法,因此是作为方法论的论文发表在《社会学方法论》(*Sociological Methodology*)期刊上的。两年以后,哈佛大学的温希普和摩根在《美国社会学年鉴》(*Annual Review of Sociology*)上系统梳理了如何通过统计方法进行因果推论,其中也谈到了倾向值匹配(Winship & Morgan, 1999)。

总体来说,在2000年之前,社会学领域内对倾向值匹配的主要关注点还是在于熟悉倾向值匹配的基本逻辑,而将其作为一种普遍适用的统计方法去研究实际问题还要到21世纪才渐渐兴起。2003年,当时还在哈佛大学学习的哈丁在《美国社会学杂志》(*American Journal of Sociology*)上发表了一篇论文研究社区贫困程度对社区少年辍学及早孕的影响,其中倾向值匹配开始像多元回归那样作为一种统计技术来验证和发展已有理论(Harding, 2003)。也正是在这一阶段,倾向值匹配的软件模块开始被开发出来并嵌入到一般的统计软件中。例如,贝克尔和栎野在2002年发布了可以嵌入Stata的用于倾向值匹配的模块“*pscore*”(Becker & Ichino, 2002)。鲁汶及夏内西也在一年后发布了他们所编写的用于倾向值匹配的模块“*psmatch2*”(Leuven & Sianesi, 2003)。这些统计模块的开发使倾向值匹配开始在社会学领域广泛应用。根据哈佛大学社会学系安卫华的梳理,截至2009年,在《美国社会学评论》(*American Sociological Review*)和《美国社会学杂志》上发表的论文中,利用倾向值匹配技术的超过了200篇,内容涉及教育社会学、互联网使用、失业问题、健康问题等诸多方面(An, 2010)。可以说,在当前的社会学研究中,倾向值匹配已逐渐发展成为一个普适的统计研究方法,并得到越来越多的重视。

倾向值匹配在社会学领域的兴起并非偶然,这是和社会学对大规模调查资料的使用分不开的。与统计学领域的实验设计不同,调查资料无法对研究个体进行人为控制。从这个意义上说,基于调查资料的研究也被称为“观察性研究”(observational studies)(Rosenbaum, 2002)。在实验性研究中,选择性误差可以通过随机化(randomization)的方法来克服,但在观察性研究中,类似的随机化是不可能的。因此,社会学家们基于调查资料去探究任何两个变量关系的时候都不可避免地会受到其他混淆变量的影响。例如,布兰德和谢宇在2010年的一篇文章中列举了在研究大学教育和收入关系时所存在的选择性误差

(Brand & Xie 2010)。一方面, 很多人之所以能够上好大学就在于其个人能力较其他人高, 而这些人在未来的工作中也很可能表现得更好从而收入更高。如果我们的研究样本大部分都是高智商群体, 那么就会“显得”大学教育带来高收入, 但实际上, 大学教育对收入的影响是被个人能力这个变量给“膨胀起来”(inflated)的。另一方面, 很多人上大学概率低是由于他们家庭贫难以负担教育费用。对这些人, 只要他们完成了大学教育, 即使毕业后的收入只是达到平均水平, 也已经比之前有很大飞跃。在这种情况下如果我们的样本偏重于家境贫寒的人, 我们就会发现大学教育再一次“显得”对收入提升有很大促进作用, 而此时的混淆变量就是家庭经济背景。对于这种选择性误差, 我们可以通过赫克曼误差修正模型(selection correction model)来进行控制, 但这个模型实施起来较为复杂。相比而言, 倾向值匹配为我们提供了一种更为直观简便的处理选择性误差的方法。在下一节, 我将通过一个实例来展示如何进行倾向值匹配。

二、倾向值匹配的实施过程: 一个实例

在这一部分, 笔者将利用倾向值匹配的方法分析中国综合社会调查 2005 年城市问卷(CGSS2005)的相关变量, 并由此来估计中国城市居民大学教育的收入回报。其中因变量是收入水平(2004 年全年总收入, 包括工资、各种奖金、补贴、分红、股息、保险、退休金、经营性纯收入、银行利息、馈赠等所有收入), 这里将其取自然对数。我们关心的自变量是被访者教育水平(1 = 大学本科教育及以上; 0 = 大学本科教育以下)。我们需要控制的混淆变量包括: 性别(1 = 女性; 0 = 男性)、年龄、政治身份(1 = 党员; 0 = 非党员)、父亲教育水平(1 = 高中教育及以上; 0 = 高中教育以下)、户口类型(1 = 城市户口; 0 = 非城市户口)、被访者的单位性质及父亲的单位性质(1 = 党政机关; 2 = 国有企业; 3 = 国有事业; 4 = 集体企业; 5 = 其他)。此处选取这些变量是因为已有研究表明这些变量会混淆教育水平和收入之间的关系(Li & Walder 2001; Lin & Bian, 1991; Shu, 2004; Walder & Hu, 2009; Xie & Wu 2008)。

一般我们会把这些混淆变量作为控制变量纳入到回归模型中, 但

这样做有一些潜在风险。首先,如果将这些混淆变量作为自变量放进回归模型,我们就潜在假定了这些混淆变量对收入的效果与大学教育对收入的效果之间存在一种线性关系。然而这种线性假定缺乏理论和实践依据(Morgan 2001)。也就是说,我们不能简单认为性别对收入的影响与教育对收入的影响是累加的(additive)。其次,大学教育的回归系数代表的是一种“平均”效果。这个系数所回答的问题是:在人口中任意选取一个人,如果他接受的是大学教育,他的收入会是什么水平。然而,在探索因果关系时我们所关心的问题则是:(1)一个任意选取的大学生如果一开始没上大学的话会是什么收入水平;(2)一个任意选取的非大学生如果上大学的话会是什么收入水平。(1)和(2)是两个不同的问题,而回归模型则没有区分它们,只是取了它们的平均水平,这样做无疑会带来误差。最后,由于混淆变量与我们关心的自变量之间存在相关性,简单地将混淆变量纳入多元回归模型有可能产生共线性问题。倾向值匹配有效地控制了这些混淆变量,同时还通过一种半非参数性(semi non-parametric)方法避免了上面提到的三个风险。^①具体而言,运用倾向值匹配方法有以下几步。

步骤一:预测倾向值。这一步是利用已知的混淆变量使用 Logistic 或 Probit 模型来预测个体进入大学读书的概率。基于上面变量,得到的结果见表 1。^②

通过表 1 我们可以发现,这些混淆变量加在一起对是否进入大学学习有比较强的解释力,这一点可以从虚拟 R^2 (Pseudo R^2) 的数值(超过 17%)看出来。对社会学研究而言,这个虚拟 R^2 值并不低,由此可以看出该模型中的混淆变量能够比较显著地预测个体能否进入大学读书。^③

步骤二:基于倾向值进行匹配。有了上面的 Probit 模型,我们就能够预测每个研究个体的倾向值。需要指出的是,虽然每个个体都有倾向值得分,但有些人的倾向值太高或太低,因此无法找到相匹配的个

① 之所以说是半参数,是因为在估算倾向值时,我们使用了参数模型(Logistic 或者 Probit),但后来估计因果效果时则没有使用参数模型,即简单比较组间差异。

② 其中父亲的教育水平因为共线性问题而被自动删除。此外,由于 CGSS2005 没有测量个体智力水平的变量,个人智力水平也未能考虑进来。

③ 这里我们并不关心某个特定自变量的效果。实际上,由于潜在的共线性问题,这里的回归系数很可能不准确。

表 1 预测倾向值的 Probit 回归结果

	回归系数	标准误	Z 值
城市户口	-1.35	.35	-3.83 ***
单位性质: 党政机关	1.89	.28	6.75 ***
单位性质: 国有企业	.11	.22	.48
单位性质: 国有事业	1.40	.22	6.26 ***
单位性质: 集体企事业	.32	.31	1.02
父亲单位性质: 党政机关	-.03	.28	-.11
父亲单位性质: 国有事业	.28	.19	1.47
父亲单位性质: 集体企事业	-.26	.31	-.84
女性	.07	.15	.44
党员	-1.38	.18	-7.66 ***
年龄	-.22	.03	-8.19 ***
年龄平方	0	0	6.81 ***
截距	6.13	.78	7.90 ***
Log likelihood = -678.365			
Pseudo R ² = 0.1746			

注: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (两端检验)。

体。这些倾向值取值非常“极端”的人因为没有与之匹配的个体存在往往无法为我们提供有用的信息,所以在随后的分析中也就没有被考虑进去,这就使得最后我们在分析高等教育的收入回报时所采用的样本量可能要比一开始的样本量小。在这个最后使用的“匹配样本”中,我们能够确保受过高等教育的个体和没有受过高等教育的个体匹配起来。“匹配样本”中倾向值的取值范围被称为“共同区间”(common support)。本研究中共同区间内的样本量为 3164(其中上过大学的为 229 人)。在这个共同区间内,我们有不同的匹配方法将受过大学教育的 229 人和没受过大学教育的 2935 人配对。对某个上过大学的个体 A,比较常用的方法包括邻近匹配(找与 A 的倾向值得分最接近的未上大学的个体 B 匹配),半径匹配(以个体 A 的倾向值为中心,以某个数值为半径,在这个范围内的所有没上过大学的个体与 A 匹配)以及核心匹配。^①

① 核心匹配比较复杂,其基本原理是将没有受过大学教育的人的收入值(因变量取值)加权平均起来,而权重则是核心方程(kernel function)的取值(详见 Guo & Fraser 2010)。

步骤三: 基于匹配样本进行因果系数估计。在这个匹配好的样本中, 我们只需比较那些上过大学和没上大学的个体的平均收入差值就可以估计出大学教育本身对收入的影响。由于配对样本的倾向值近似, 配对个体在混淆变量上的取值极为近似, 这样也就控制了混淆变量的影响。换句话说, 我们得到的组间(上大学和没上大学) 差异就只能归因于大学教育的有无。我们最后的结果如表 2 所示。

表 2 倾向值匹配的结果

	受过大学教育的人	没受过大学教育的人	因果系数	标准误	T 值
邻近匹配	233	2971	. 730	. 057	12. 87 ***
半径匹配(半径 0. 01)	230	2971	. 770	. 055	13. 93 ***
核心匹配	232	2971	. 764	. 053	14. 36 ***

注: (1) 由于我们这里主要是用没有受过大学教育的人去匹配受过大学教育的人, 我们的关注点是那些受过高等教育的人(即接受了某种“处理”的人) , 因此这里的因果系数即“受到处理的个体的平均处理效果”(average treatment effect of the treated) , 简称为 ATT; (2) * $p < 0. 1$, ** $p < 0. 05$, *** $p < 0. 01$ (两端检验) 。

通过表 2 可以发现, 不同匹配方法下我们得到的因果系数不完全相等, 这是因为受过大学教育的个体在不同的匹配方式下可能和不同的未受大学教育的个体进行了匹配。然而这些不同的匹配方法得出的结论基本一致(系数在 0. 73 到 0. 77 之间) 。这些系数就是大学教育的收入回报。其可以解释为与没有受过大学教育的人相比, 大学教育带来 $\log(\text{年收入})$ 的变化。

需要说明的是, 在社会学定量研究中, 在完成了匹配以后, 通常会比较每个混淆变量在大学组与非大学组之间是否还存在显著差异。理想情况是大学组与非大学组的混淆变量均值没有显著差异, 这时我们可以说这样的配对样本是“平衡”(balanced) 的。限于篇幅, 这里未将均值比较结果展示出来, 但我们上面的分析已通过了平衡性检验。此外, 倾向值匹配通常配合敏感性分析(sensitivity analysis) 来检验我们在预测倾向值时是否忽略了某些关键的混淆变量。在敏感性分析阶段, 我们假设存在一个或多个我们没有能够控制的混淆变量, 如果这些假设的未控制变量对收入的影响的大范围变动都无法改变我们的结论,

我们的结论就站得住脚。上面的研究就通过了相关的检验。^①

三、倾向值匹配与因果推论

倾向值匹配后的结果不仅仅指出了变量之间有联系,还进一步确立了二者之间的因果性。这种方法论上的优势可以从科技哲学和统计学两个方面予以阐释。

(一) 科技哲学角度的阐释

倾向值匹配对因果性结论的支持首先依据的是科技哲学领域内对因果关系的理论探讨。当我们谈及因果性的时候,我们往往会回归到密尔对求同法(method of agreement)及求异法(method of differences)的讨论(彭玉生,2011; Mill,2002/1984; Sobel,1995,1996)。求同法是指在一个群体中所有人都在两个变量上取值相同(例如所有人都上了大学,同时所有人都是高智商)而在其他变量上取值不同(他们不全是女性,不全是农村居民,等等),那么这两个变量之间(大学教育和智力水平)就具有因果关系。求异法是指两个个体在因变量上的取值不同,而在某个自变量之外的其他自变量上的取值相同,则那个取值不同的自变量和因变量之间存在因果关系。正如彭玉生所论述的那样,密尔的这两种探索因果关系的方法对后来的统计实验设计有着巨大影响(彭玉生,2011)。无论是求同法还是求异法,二者都要求对关键自变量之外的其他自变量进行考察。只有其他自变量满足特定的条件时(求同法要求其他自变量的取值都不一样,而求异法要求其他自变量的取值都一样)我们才能够确定某种因果关系。这实际上是引入了“控制”的思想,即只有“控制”了其他变量,我们才能够真正确定我们

^① 在敏感性分析中,比较常用的方法是罗森鲍姆界限(Rosenbaum Bounds)。在该方法中,我们用伽玛系数来指代被忽视的混淆变量对人们是否进入大学读书产生的影响。使用罗森鲍姆界限进行敏感性分析时,我们主要观察在什么样的伽玛系数水平上我们的结论变得不显著。如果在伽玛系数接近1的时候我们的结论就已经不显著了,那么我们已有的结论就经不起推敲。但是如果在伽玛系数取值很大(通常接近2)的时候我们的已有结论才变得不显著,我们就能够认为我们的结论是站得住脚的(Lin et al.,1997; Rosenbaum & Rubin,1983)。本实例中我们的结论直到伽玛系数在3.5时才在0.05的水平上不显著。

关心的两个变量之间存在因果关系。

从“控制”的角度出发,倾向值匹配十分巧妙地完成了对多个混淆变量的控制。我们可以从一个混淆变量的情况谈起:假定只有个人智力水平混淆了大学教育和收入之间的关系。一个比较直观的控制个人智力的办法是将个人智力这一变量细分(sub-classification)成不同层次以保证每一个层次中的人的个人智力水平近似。然后我们在各个层次内部观察大学教育和收入的关系,最后将这些关系综合起来(Rubin, 1997)。如果存在两个混淆变量(例如家庭经济背景好坏和个人智力高低)我们可以将这两个变量交互分成 2×2 个小组,组内个体在这两个变量上的取值都是一样的(家庭背景好且智力高、家庭背景好且智力低、家庭背景不好且智力高,以及家庭背景不好且智力低),在每组内部观察大学教育和收入的关系然后综合起来。至此,通过细分的方法,我们完成了“控制”混淆变量的工作。但随着混淆变量越来越多,这种细分法就变得十分不方便了。例如,如果我们要控制5个混淆变量,每个变量有5个取值水平,我们就需要划分 $5^5 = 3125$ 个小组。很快我们的数据样本量就不够保证每一组都有个体。倾向值匹配的办法巧妙地解决了这个多混淆变量下的“多维”(multiple dimensional)问题:它不再关注每个需要控制的混淆变量的具体取值,而是转而关注将这些变量纳入Logistic回归方程后预测出来的倾向值取值。只要保证倾向值匹配,这些所有需要控制的混淆变量就都考虑到了(Rosenbaum & Rubin, 1983)。这样做实际上是将对多个混淆变量的控制转为对倾向值的控制从而达到“降维”(dimension reduction)的目的。换句话说,无论有多少需要控制的混淆变量,我们都能够通过倾向值匹配的方法将它们控制,从而帮助我们得出因果性结论。因此,从“控制”的角度出发,倾向值匹配法很好地解决了多混淆变量时的控制问题,从而支持了因果推论。^①

(二) 统计学角度的阐释

倾向值匹配的因果推论功能也能通过统计学视角进行阐释。这里

① 倾向值匹配的另一个优势在于,如果我们有很多需要控制的混淆变量,它们之间有可能互相抵消或者强化。如果使用细分的方法去考察单个变量,这种抵消或强化效果就无法体现出来。相反,倾向值匹配预测出的倾向值通过回归的方式考察了这种互相抵消或强化作用。

我们需要引入反事实框架(counterfactual framework) 这一统计学理论。反事实(counter facts) 是指相反情境下的某种状态。例如, 一群病人在一个实验中被分到实验组接受新药物治疗。这些人症状的减轻或加重是我们能够观察到的“事实”。而“反事实”则是指“假设”这一群病人当时不是被分到实验组而是对照组, 即没有接受新药物治疗, 那么他们的症状会是什么样子。自变量(新药) 对于症状的因果性效果在统计学意义上就是指这“同一群人”在实验组时的症状和在对照组时的症状之间的差异。换句话说, 统计学上的因果关系是可观察到的“事实”与其“反事实”之间的差异。从反事实的框架出发, 因果性的关系可以表示为:

$$\begin{aligned}
 T &= \pi [E(Y_1 | w = 1) - E(Y_0 | w = 1)] + \\
 &\quad (1 - \pi) [E(Y_1 | w = 0) - E(Y_0 | w = 0)] \\
 &= [\pi E(Y_1 | w = 1) + (1 - \pi) (E(Y_1 | w = 0))] - \\
 &\quad [\pi E(Y_0 | w = 1) + (1 - \pi) E(Y_0 | w = 0)]
 \end{aligned}$$

(反事实框架下的因果推论公式)

在这个公式里, T 是指因果关系; π 是指所有调查对象中在实验组中的比例, 而 $1 - \pi$ 也就表示了所有调查对象中在对照组的比例(例如 $1/3$ 是大学生, 那么 $\pi = 1/3$, 而没上大学的人的比例就是 $1 - \pi = 2/3$); w 是一个二变变量, 其中 1 代表个体在实验组而 0 代表个体在对照组; Y_1 和 Y_0 分别指代实验组和对照组的成员在因变量上的取值。 E 则是取平均值的意思。在上述公式中, $E(Y_1 | w = 1)$ 或 $E(Y_0 | w = 0)$ 是可观测到的事实, 而 $E(Y_1 | w = 0)$ 和 $E(Y_0 | w = 1)$ 则是反事实。^① 因果关系 T 就表示为实验组中的个体其“事实”与“反事实”之间的差异——即 $E(Y_1 | w = 1) - E(Y_0 | w = 1)$ ——与对照组中的个体其“事实”与“反事实”之间的差异——即 $E(Y_1 | w = 0) - E(Y_0 | w = 0)$ ——的加权平均值(权重分别为 π 和 $1 - \pi$)。

但问题在于, 我们永远也不可能观测到反事实是什么, 因为在某项特定的研究中, 某一群人只可能在实验组或对照组, 而不能同时在两组中出现。这被称为“因果推论的基本问题”(Holland, 1986)。为了做出

^① 这里, $E(Y_1 | w = 0)$ 是指在对照组中的人“如果”被放置在实验组时所得到的因变量均值。而 $E(Y_0 | w = 1)$ 则是指在实验组中的人“如果”被放置在对照组内的话所得到的因变量均值。

因果推论,我们需要用可观测到的 $E(Y_1 | w = 1)$ 和 $E(Y_0 | w = 0)$ 来简化上面的因果推论公式。具体而言,我们希望能够满足以下条件,这在统计学上称为“非混淆假设”(unconfoundedness assumption):

$$E(Y_1 | w = 0) = E(Y_1 | w = 1) \text{ 以及}$$

$$E(Y_0 | w = 0) = E(Y_0 | w = 1) \quad (\text{非混淆假设})$$

不难看出,如果满足非混淆假设,反事实框架下的因果推论公式就简化为:

$$T = E(Y_1 | w = 1) - E(Y_0 | w = 0) \quad (\text{简化因果推论公式})$$

如果简化因果推论公式右边的两项均能观测到,我们就能做出因果性结论,因此,从反事实框架出发,能否做出因果推论主要取决于非混淆假设是否满足。^① 随机化是统计学中常用的办法,由于实验个体是通过随机方式分配到实验组和对照组中的, w 本身就 and 最后的实验结果 Y_1 或 Y_0 没有关系了。^② 换句话说,无论 $w = 0$ 还是 $w = 1$, $E(Y_1)$ 或 $E(Y_0)$ 的值都是固定的,进而 $E(Y_1 | w = 0) = E(Y_1 | w = 1)$ 且 $E(Y_0 | w = 0) = E(Y_0 | w = 1)$ 。但是,对基于调查资料的社会学研究来讲,我们做不到随机化。为了满足非混淆假设,我们能做的是尽可能控制混淆变量,并保证这些变量一旦被控制起来, w 就能够近似地和 Y_1 或 Y_0 保持独立,换句话说,我们希望做到:

$$E(Y_1 | w = 0, X) = E(Y_1 | w = 1, X) \text{ 以及}$$

$$E(Y_0 | w = 0, X) = E(Y_0 | w = 1, X) \quad (\text{非混淆假设: 控制变量})$$

上式表明,只要能够找到并控制混淆变量 X , 我们就能够近似地做到 w 独立于 Y 。在倾向值匹配中,所有的这些 X 通过 Logistic 回归总结成为一个特定的倾向值 P ,而非混淆假设就是通过控制倾向值 P 来满足的。即:

$$E(Y_1 | w = 0, P) = E(Y_1 | w = 1, P) \text{ 以及}$$

$$E(Y_0 | w = 0, P) = E(Y_0 | w = 1, P)$$

(非混淆假设: 控制倾向值)

至此,通过控制倾向值,我们可以“近似地”满足统计学反事实框

① 当然,在反事实框架下的因果推论还预设了其他条件,比如个体之间没有相互影响(stable unit treatment value assumption)等。这些预设的前提条件会在下面的讨论中提到(具体可以参阅 Guo & Fraser 2010; Morgan & Winship 2007)。

② 根据大数定理,个体的各种不可控差异,在随机分配过程中被平均了。换句话说,两组人可以近似地被认为“总体上同质”。

架下的非混淆假设从而做出因果推论。之所以说“近似”,是因为控制倾向值的效果毕竟不是真正的随机化。很多时候我们很难知道是否已经控制了需要控制的“所有”混淆变量。正因为如此,在完成倾向值匹配以后我们通常需要进行敏感性分析。在这个意义上说,倾向值匹配只是努力地满足反事实框架下的因果推论条件,但即使如此,我们也已经从统计学意义上理解了倾向值匹配为什么能够帮助我们得出因果性结论。

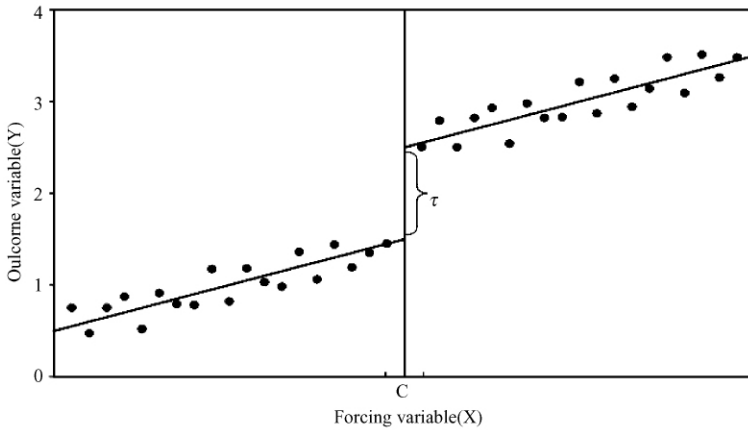
四、倾向值匹配与其他社会科学方法的比较

以上分析从科技哲学及统计学的角度阐释了为什么通过倾向值匹配可以得到因果性结论。然而,因果性分析的方法并不局限于倾向值匹配,在教育学及计量经济学领域中有很多其他的方法可以帮助我们建构因果关系。在以下讨论中,我将把倾向值匹配与回归中断设计和工具变量进行比较以展示倾向值匹配的优势。

(一) 倾向值匹配和回归中断设计

在教育学中应用比较广泛的因果推论方法是“回归中断设计”(regression discontinuity design)。该方法最先由两位美国学者在1960(Thistlethwaite & Campbell, 1960)提出。在他们的研究中,两位学者关心的是学习上的荣誉奖励(原因)是否能够提升学生未来的学术成就(结果)。这里的荣誉奖励是根据考试成绩而定的:当考试成绩 x 超过一定分数 c ,则给予奖励($D=1$),否则($x < c$ 时)则没有奖励($D=0$)。通过这种操作,我们就在我们所关心的自变量那里建立了一种“中断”(c 之上和之下)。随后如果学生的学术成就也发生了类似的中断(例如考试成绩在 c 以下学生的学术成就低于考试成绩在 c 以上的学生的学术成就),则可以认为奖励和学术成就之间有因果关系。图1用图示的方式表达了这种关系。

在图1中, x 在 c 处的中断对应于因变量 Y 的中断,这种中断 τ 代表了 x 对于 Y 的效果。从反事实框架来看,回归中断设计满足了非混淆假设。在上面的荣誉奖励例子中,一个人得到荣誉奖励可以理解为一个人进入了实验组,而这“完全”取决于学生的考试分数 x (分数高就



注: 横坐标 (X) 代表考试成绩 纵坐标 (Y) 代表学术成就。

图片来源: Lee & Lemieux ,2009。

图1 回归中断设计的基本逻辑

有奖励而分数低则没有奖励)。因此,只要我们控制了分数 x , w (是否得到奖励) 就和因变量 Y_1 (得到奖励的学生的学术成就) 以及 Y_0 (没有得到奖励的学生的学术成就) 独立了。用公式可以表述如下:

$$E(Y_1 | w = 0, x) = E(Y_1 | w = 1, x) \text{ 以及}$$

$$E(Y_0 | w = 0, x) = E(Y_0 | w = 1, x)$$

(非混淆假设: 回归中断设计)

这里我们可以将回归中断设计与倾向值匹配进行一下比较。倾向值匹配将对多个混淆变量的控制转化为对“单一倾向值”的控制,以此满足非混淆假设。而回归中断设计则是通过控制一个变量(如上例中的考试成绩)就能完成这一点。回归中断设计之所以不需要考虑多个混淆变量就在于一个个体是否能够接受某个自变量的影响(即能否得到奖励)完全取决于单一变量 x (考试成绩)。换句话说,回归中断设计中只有一个混淆变量 x 。然而,在社会学的调查研究中很难找到这种“单一混淆变量”的情况。正因为如此,回归中断设计在社会学中的应用远不及在教育学中普遍(Angrist & Lavy, 1999; Black, 1999)。此外,回归中断设计的另一个缺陷在于:如果存在其他自变量也出现某种“中断”的情况,我们就很难知道什么原因造成了因变量取值的“中断”。假设一个人是否获奖不仅仅与考试成绩在临界点上下有关($x > c$ 或

$x < c$) ,也和年龄有关(例如年龄大的人容易获奖,而年龄小的人很难获奖) ,则回归中断设计就无法探究荣誉奖励本身对未来学术成就的影响(因为学术成就的中断也有可能是因为年龄不同) (Hahn et al. , 2001) 。在社会学研究中,类似的多重中断并不少见,这也使得回归中断设计的应用范围受到限制。

(二) 倾向值匹配和工具变量

除了回归中断设计,在经济学及传染病学领域应用比较广泛的另外一个方法是工具变量(instrumental variable) (Angrist et al. ,1996; Angrist & Krueger 2001; Greenland 2000; Martens et al. 2006; Newhouse & McClellan ,1998) 。假定我们希望探索变量 X 对 Y 的因果关系,混淆变量 U 的存在会影响我们的研究。这时我们寻找到一个工具变量 Z,该变量的特点在于和 X 相关,但不和 Y 直接相关(除非是通过 X) ,同时 Z 也不与混淆变量 U 相关。这样,我们就不直接考察 X 是否能影响 Y ,而是转而考察 Z 对 Y 的影响。如果 Z 很明显地影响了 Y ,我们就能得出 X 影响了 Y 的结论。工具变量的基本逻辑表示如图 2。

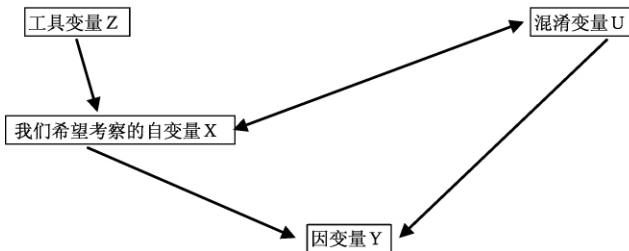


图 2 工具变量的基本逻辑

通过图 2 我们很容易理解工具变量的基本逻辑,由于混淆变量 U 和 X 互相影响,我们很难直接探索 X 对 Y 的“纯”作用。所以我们找到工具变量 Z。Z 和 Y 的关系只能通过 X,因此如果我们能够发现 Z 对 Y 有作用,我们就能间接得出结论说 X 对 Y 有作用,从而确定了一种因果关系。

从严格的统计学意义上来说,利用工具变量进行因果推论需要满足很多假设条件(Imbens & Angrist,1994) 。这里可以通过一个经典实

例来展示这些假设。安格里斯特等人(Angrist et al. ,1996) 使用工具变量分析了服兵役(X) 是否会对士兵在退伍后的健康状况(Y) 产生因果性影响。这里 服兵役的工具变量是抽签的结果(Z) 。当抽到的数字低于某个值则服兵役 ,而高于这个值则不需服兵役。在这个例子中 ,安格里斯特及其同事阐述了使用工具变量进行因果推论的几个关键假设。第一个假设称为“单位实验效果值稳定假设”(stable unit treatment value assumption) 。这意味着每个人的抽签结果不会对他人是否服兵役产生影响。^① 第二个假设就是上面提到的非混淆假设 ,即抽签的结果和个体的健康之间相互独立。由于这里是随机抽签 ,因此这一条件得到满足。第三个假设是“排除性假设”(exclusion restriction) 。这一假设意味着在因果推论中排除两类人。第一类人是无论抽签结果如何都会服兵役(总是服兵役的人) ,而第二类人是无论抽签结果如何都不会服兵役(总是不服兵役的人) 。^② 第四个假设是“平均因果效果非零”(nonzero average causal effect) ,即保证如果抽中的号码小 ,则服兵役的概率就高。最后一个假设是“单调性”(monotonicity) ,意味着抽中小数字的人去服兵役的概率要比抽中大数字的人去服兵役的概率高。^③ 根据这一假设 ,那些“低数字偏偏不服兵役”而“高数字偏偏服兵役”的“叛逆者”就被排除在外了。^④

根据这些假设 ,我们回到反事实框架下的因果推论公式 ,即:

$$T = [\pi * E(Y_1 | w = 1) + (1 - \pi) * E(Y_1 | w = 0)] - [\pi * E(Y_0 | w = 1) + (1 - \pi) * E(Y_0 | w = 0)]$$

这里由于我们有抽签的机制 , π (即服兵役的比例) 就等于 $X_{(Z=1)}$,而没有服兵役的比例 $1 - \pi$ 则等于 $X_{(Z=0)}$,而 $w = X$ 。其中 , X 表示是否服兵役 ,它是 Z 的函数 ,而 Z 表示抽签结果($Z = 1$ 表示抽中小号 ,而 $Z = 0$ 表示抽中大号) 。将它们代入反事实框架下的因果推论公式 ,我们

① 数学表达式为: 如果 $Z = Z'$,那么 $X_{(Z)} = X_{(Z')}$; 如果 $Z = Z'$ 而且 $X = X'$,那么 $Y(Z, X) = Y(Z', X')$ 。这里 Z 表示抽签号 ,而 X 表示是否服兵役。因为服兵役是根据抽签号决定 ,所以 X 可以表示成 Z 的函数 ,即 $X_{(Z)}$ 。

② 排除性假设认为对于这两类人 ,他们的健康结果都和其他通过抽签而服兵役或不服兵役的人的健康状况相同。其数学表达式为: $Y(Z, X) = Y(Z', X) = Y(X)$,也就是说 ,最后的健康状况是和服兵役与否而不是和抽签号本身直接挂钩的。

③ 数学表达式为: $P [X_{(Z=1)}] > P [X_{(Z=0)}]$ 。

④ 这些人在安格里斯特等人的文章中被称作“叛逆者”(defier) ,因为这里的规定是小数字服兵役而大数字不服兵役 ,但他们总是做“相反”的事情 ,即抽到小数字偏偏不服兵役 ,而抽到大数字却偏偏去服兵役。这些人在因果推论中不予考虑。

得到:

$$T = [X_{(Z=1)} * E(Y_1 | X = 1) + X_{(Z=0)} * (E(Y_1 | X = 0) - [X_{(Z=1)} * E(Y_0 | X = 1) + X_{(Z=0)} * E(Y_0 | X = 0)])]$$

(因果推论公式: 工具变量)

由于叛逆者被排除出去, 可以假定 $E(Y_0 | X = 1) = E(Y_0 | X = 0) = 0$, 这进一步将因果关系简化成为:

$$T = X_{(Z=1)} E(Y_1 | X = 1) - X_{(Z=0)} E(Y_0 | X = 0)$$

(简化因果推论公式: 工具变量)

上述公式右端的各项都是可观测到的, 从而使得因果推论得以完成。然而, 使用工具变量方法存在诸多限制。首先, 我们所推论出的因果关系被称为“局部平均治疗效果”(local average treatment effect) (Wooldridge 2002)。这是因为我们最终能够推算出的因果关系仅仅适用于那些遵循“拿到小号服兵役, 拿到大号不服兵役”原则的“服从者”(complier) (Imbens & Angrist, 1994), 但我们不能考察总是服兵役的人、总是不服兵役的人或叛逆者。其次, 工具变量方法能否带来因果推论很大程度上取决于我们能否找到好的工具变量。如果工具变量本身不够好(比如工具变量和我们关心的自变量的联系很弱, 或者和某些混淆变量有关系), 那么我们的结论就站不住脚了(Bound et al., 1995)。最后, 工具变量的实施需要满足一定的随机性以满足非混淆假设。但在社会学研究中, 类似于随机抽签这样“完美”的工具变量几乎很少见到。

总之, 无论是回归中断设计还是工具变量都不能够像倾向值匹配那样可以方便地应用于基于调查研究的数据分析。回归中断设计需要寻找一个变量“完全决定”个体是否进入实验组或对照组, 而工具变量的方法也需要寻找到合适的工具变量来满足上面列举的诸多条件, 这些在调查数据中往往难以满足。正因为如此, 在社会学研究中倾向值匹配的方法要比其他方法更为“流行”。

五、倾向值匹配的限制性

任何统计方法都有其自身的局限性, 倾向值匹配也不例外。首先, 在社会学研究中, 非混淆假定往往难以满足。以本文开始的实例来说,

受问卷变量的限制,我们很可能没有办法控制一些影响人们上大学的因素(例如个人智力水平)。正因为如此,很多研究会明确说明其结论有效性的前提是非混淆假设成立,如果假设不成立,结论也会相应受到质疑(Brand & Xie 2010)。

其次,倾向值匹配假定个体之间不存在相互影响,即“单位实验效果值稳定假设”。该假设也同样会在社会学研究中受到质疑。举例而言,摩根(Morgan, 2001) 在研究天主教学校对学生成绩的影响时指出,“单位实验效果值稳定”意味着天主教学校对个体 A 的影响与其对个体 B 的影响无关,照此推论,因为彼此互不影响,接触天主教学校教育的人数多少就不是一个需要考虑的问题。换句话说,天主教学校教育对学生成绩的影响和学校人数无关。无疑这和现实情况相左。因此,在利用倾向值匹配去研究组织结构时,研究者需要十分注意这一假设。

第三,除上述基本假设之外,倾向值匹配并不考虑变量之间的交互作用而只是关注某个我们关心的自变量的效果,同时和多元回归一样,倾向值匹配也会受到忽略变量(omitted variables) 问题的影响。此外,目前主流统计软件分析的自变量是二分变量(例如上不上大学),而对连续型自变量的探讨还很不成熟。

总体而言,倾向值匹配方法主要应用于效果评估研究中(evaluation)。例如上述大学教育对收入的影响以及摩根研究的天主教学校对学生成绩的影响等等。这类问题的一个共同点在于我们在分析数据之前已经预先知道一个二分型的变量对于另外一个变量有因果效应。此外,对那些影响二分变量的混淆变量,我们也应该具有比较清楚的理论支持。在这种情况下,倾向值匹配会比较适用。然而,如果我们对谁“影响”谁不是十分清晰(例如互为因果),或者对于我们需要考虑哪些混淆变量也不了解的话,倾向值匹配就很有可能违反上面提及的前提假设,从而使我们的结论受到质疑。

六、总结及讨论

社会学研究中对因果性的重视使我们日益关注并应用倾向值匹配方法。本文系统梳理了倾向值匹配的历史、发展及其对社会学调查研究的意义,并通过一个具体实例展示了倾向值匹配的实施过程。此外,

本文从科技哲学理论层面与统计学数理层面论述了为什么倾向值匹配可以帮助我们进行因果推论。然后, 本文对比了倾向值匹配与回归中断设计及工具变量方法在因果推论中的应用。最后本文简要陈述了倾向值匹配方法的局限性和适用性。

作为一种新兴统计技术, 倾向值匹配还在不断发展。首先, 在最近几年中, 很多新的匹配技术被开发出来。例如完全匹配方法(full matching) 可以提高样本的利用率从而保证每个个体都能在匹配过程中被使用到(Hansen, 2004; Stuart & Green, 2008)。遗传性匹配(genetic matching) 则利用人工智能的方法提高匹配效率, 其过程甚至不需要计算具体的倾向值(Sekhon & Mebane, 1998)。在未来一段时间内, 如何开发新的匹配技术将是统计学家及社会科学家共同努力的方向。其次, 倾向值匹配方法的另一个发展趋势在于对更多研究情景的考虑。例如, 新的技术开始被开发出来以考察如何将倾向值匹配应用于自变量是连续型的情况(Hirano & Imbens, 2004) 或自变量超过两个取值水平的情况(Lechner & Miquel, 2005)。最后, 倾向值匹配和现有统计模型和统计技术也越来越多地结合起来。这方面已有的探索包括将倾向值匹配和分层模型结合起来考察因果效果在不同群体中体现出的异质性(Brand, 2010; Brand & Davis, 2011; Brand & Xie, 2010), 以及利用贝叶斯方法考察倾向值计算的不确定性(An, 2010)。这三个发展趋势预示着倾向值匹配方法会在未来一段时间有更大的飞跃。

参考文献:

彭玉生, 2011, 《社会科学中的因果分析》, 《社会学研究》第3期。

An, Weihua 2010, "Bayesian Propensity Score Estimators: Incorporating Uncertainties in Propensity Scores into Causal Inference." *Sociological Methodology* 40(1).

Angrist, Joshua D. & Alan B. Krueger 2001, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4).

Angrist, Joshua D. & Victor Lavy 1999, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2).

Angrist, Joshua D., Guido W. Imbens & Donald B. Rubin 1996, "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434).

Becker Sascha O. & Andrea Ichino 2002, "Estimation of Average Treatment Effects Based on Propensity Scores." *The Stata Journal* 2(4).

Black, Sandra E. 1999, "Do 'Better' Schools Matter? Parental Valuation of Elementary Educa-

- tion.” *Quarterly Journal of Economics* 114.
- Bound, John, David A. Jaeger & Regina M. Baker 1995, “Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak.” *Journal of the American Statistical Association* 90(430) .
- Brand, Jennie E. & Yu Xie 2010, “Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education.” *American Sociological Review* 75(2) .
- Brand, Jennie E. 2010, “Civic Returns to Higher Education: A Note on Heterogeneous Effects.” *Social Forces* 89(2) .
- Brand, Jennie E. & Dwight Davis 2011, “The Impact of College Education on Fertility: Evidence for Heterogeneous Effects.” *Demography* 48(3) .
- Greenland, Sander 2000, “An Introduction to Instrumental Variables for Epidemiologists.” *International Journal of Epidemiology* 29(4) .
- Guo, Shenyang & Mark W. Fraser 2010, *Propensity Score Analysis: Statistical Methods and Applications*. CA: Sage Publications.
- Hahn, Jinyong, Petra Todd & Wilbert Van der Klaauw 2001, “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica* 69(1) .
- Hansen, Ben B. 2004, “Full Matching in an Observational Study of Coaching for the SAT.” *Journal of the American Statistical Association* 99(467) .
- Harding, David J. 2003, “Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy.” *American Journal of Sociology* 109(3) .
- Heckman, James J. & Jeffrey A. Smith 1995, “Assessing the Case for Randomized Social Experiments.” *Journal of Economic Perspectives* 9(2) .
- Heckman, James J. 1978, “Dummy Endogenous Variables in a Simultaneous Equations System.” *Econometrica* 46.
- 1979, “Selection Bias as A Specification Error.” *Econometrica* 47.
- Hirano, Keisuke & Guido W. Imbens 2004, “The Propensity Score with Continuous Treatments.” In Andrew Gelman & Xiaoli Meng (eds.) , *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. New York: Wiley.
- Holland, Paul W. 1986, “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396) .
- Imbens, Guido W. & Joshua D. Angrist 1994, “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62(2) .
- Lechner, Michael & Ruth Miquel 2005, “Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions.” Working Paper, SIAW.
- Lee, David S. & Thomas Lemieux 2009, “Regression Discontinuity Designs in Economics.” NBER Working Paper Series(Retrieved from <http://www.nber.org/papers/w14723>) .
- Leuven, Edwin & Barbara Sianesi 2003, “Psmatch 2: State Module to Perform Full Mahalanobis

- and Propensity Score Matching , Common Support Graphing , and Covariate Imbalance Testing.” (<http://ideas.repec.org/c/boc/bocode/s432001.html>)
- Li , Bobai & Andrew G. Walder 2001 , “Career Advancement as Party Patronage: Sponsored Mobility into the Chinese Administrative Elite , 1949 – 1996.” *American Journal of Sociology* 106 (5) .
- Lin , Danyu , Bruce M. Psaty & Richard A. Chronmal 1997 , “Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies.” Technical Report #144. Seattle: University of Washington School of Public Health.
- Lin , Nan & Yanjie Bian 1991 , “Getting Ahead in Urban China.” *American Journal of Sociology* 97 (3) .
- Martens , Edwin P. , Wiebe R. Pestman , Anthonius de Boer , Svetlana V. Belitser & Olaf H. Klun-
gel 2006 , “Instrumental Variables Application and Limitations.” *Epidemiology* 17(3) .
- Mill , John Stuart 2002/1984 , *System of Logic: Ratiocinative and Inductive*. Hawaii: University Press of the Pacific.
- Morgan , Stephen L. & Christopher Winship 2007 , *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Morgan , Stephen L. & David J. Harding 2006 , “Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice.” *Sociological Methods & Research* 35(1) .
- Morgan , Stephen L. 2001 , “Counterfactuals , Causal Effect Heterogeneity , and the Catholic School Effect on Learning.” *Sociology of Education* 74.
- Newhouse , Joseph P. & Mark McClellan 1998 , “Econometrics in Outcomes Research: The Use of Instrumental Variables.” *Annual Review of Public Health* 19.
- Normand , Sharon-Lise T. , Mary Beth Landrum , Edward Guadagnoli , John Z. Ayanian , Thomas J. Ryan , Paul D. Cleary & Barbara J. McNeil 2001 , “Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores.” *Journal of Clinical Epidemiology* 54(4) .
- Rosenbaum , Paul R. 2002 , *Observational Studies*. New York: Springer.
- Rosenbaum , Paul R. & Donald B. Rubin 1983 , “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1) .
- Rubin , Donald B. 1997 , “Estimating Causal Effects from Large Data Sets Using Propensity Scores.” *Annals of Internal Medicine* 127(8) .
- Sekhon Jasjeet Singh & Walter R. Mebane , Jr. 1998 , “Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models.” *Political Analysis* 7.
- Shu , Xiaoling 2004 , “Education and Gender Egalitarianism: The Case of China.” *Sociology of Education* 77(4) .
- Smith , Herbert L. 1997 , “Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies.” *Sociological Methodology* 27.
- Sobel , Michael E. 1995 , “Causal Inference in the Social and Behavioral Sciences.” In Gerhard Arminger , Clifford C. Clogg & Michael E. Sobel (eds.) , *Handbook of Statistical Modeling for*

- the Social and Behavioral Sciences*. New York: Plenum Press.
- 1996, “An Introduction to Causal Inference.” *Sociological Methods & Research* 24(3) .
- 2000, “Causal Inference in the Social Sciences.” *Journal of the American Statistical Association* 450(5) .
- Stuart , Elizabeth A. & Kerry M. Green 2008, “Using Full Matching to Estimate Causal Effects in Nonexperimental Studies: Examining the Relationship Between Adolescent Marijuana Use and Adult Outcomes.” *Developmental Psychology* 44(2) .
- Thistlethwaite , Donald L. & Donald T. Campbell 1960, “Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment.” *Journal of Educational Psychology* 51(6) .
- Walder , Andrew G. & Songhua Hu 2009, “Revolution , Reform , and Status Inheritance: Urban China , 1949 – 1996.” *American Journal of Sociology* 114(5) .
- Winship , Christopher & Michael E. Sobel 2004, “Causal Inference in Sociological Studies.” In Melissa A. Hardy & Alan Bryman (eds.) , *Handbook of Data Analysis*. CA: Sage Publications.
- Winship , Christopher & Stephen L. Morgan 1999, “The Estimation of Causal Effects from Observational Data.” *Annual Review of Sociology* 25.
- Wooldridge , Jeffrey M. 2002 , *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.
- Xie , Yu & Xiaogang Wu 2008, “Danwei Profitability and Earnings Inequality in Urban China.” *The China Quarterly* 195.

作者单位: 美国普度大学社会学系
责任编辑: 杨 典

ratism , that's to say , we should find out the ideological and societal foundations of corporatism and its institutional characteristics. Based on the comprehensive understanding of corporatism , the author points out that although the institutional forms in the relationship between state and society in China are highly similar to those of the corporatist model , the corporatist model cannot be applied in China properly due to lack of societal foundations.

“The Trajectory of Up-Down-Fragmentation-Integration”: The producing of community space *Huang Xiaoxing* 199

Abstract: The author first analyzes the applicability of the western organizational paradigm of the producing of community space in the context of the ambiguous-strategic relationship of state and society in China , and then points out that we need to construct a new theory better matched to the transition period of China. The paper focuses on the story of the planning and construction of Guangan Cancer Hospital in Nanxin Garden Community , and analyzes the mechanisms of the producing of community space under the structured context of the reframing of the urban space. The author points out that when the layout of urban space was being carried out , power was hidden in the market system and the discourse of procedure legitimacy and property rights were used to acquire the monopoly position. Simultaneously , people in the community fought the government with the consciousness of rules and rights , but failed in the end. The paper analyzes the strategies of the hegemony of space , the tactics of space protest , and the tracks of the producing of community space. Based on the analysis , the author suggests that we should shift to the producing of the space of socialism in order to acquire space justice.

REVIEW

Propensity Score Matching and Causal Inference: A methodological review *Hu Anning* 221

Abstract: Propensity score matching method has attracted more and more attention as sociological studies begin to pay closer attention to causal relationship. This article reviews the method from the following aspects. First , the author systematically reviews the history and development of propensity score matching as well as its implications for sociological survey research. Second , this research demonstrates how to perform propensity score matching using an example. Third , based on Mill's discussion on causality and statistical counterfactual framework , this article shows how propensity score matching contributes to causal conclusion by controlling for confounding variables and satisfying the unconfoundedness assumption. Fourth , this paper compares propensity score matching with regression discontinuity design and instrumental variable method. Finally , after summarizing the major limitations of propensity score matching , this article discusses the newly development of propensity score matching in recent years.