

# 社会科学中的因果解释： 逻辑、样本与方法的权衡

叶成城

**[摘要]** 实证研究在社会科学领域兴起后，比较方法成为社会科学中最为重要的研究方法之一。严谨的实证研究离不开科学的研究设计，需要依照和遵循一定的程序步骤来进行因果解释，包括对自变量和因变量的选取、理论构建和假设检验等众多步骤。因果解释需要通过形式逻辑来进行推断，密尔最早系统性地提出了五种因果推断的方法。这五种逻辑方法对应了过程追踪、定性案例比较、定性比较分析（QCA）、定量研究4种基本的案例研究方法。前面两种方法侧重于研究案例内部的观测值，后面两种方法则侧重于讨论案例之间的平均效应，它们在不同类型和数量的案例中存在各自的联系、优势和缺陷。研究者可以通过混合方法等策略来一定程度地实现扬长避短。

**[关键词]** 社会科学 密尔方法 案例比较 因果解释 方法权衡 定量定性研究

## 一、社会科学方法论：两种错误的观点

在社会科学中，方法论的争论是如此普遍而纷杂，尤其是政治学界普遍认为存在两种完全不同的范式，即定性方法论范式和定量方法论范式，二者之间存在激烈的争论。这类方法

---

作者简介：叶成城（1987—），上海社会科学院国际问题研究所副研究员，200062。

论的争论很大程度上来自于加里·金(Gary King)、罗伯特·基欧汉(Robert Keohane)和悉尼·维巴(Sidney Verba)所合著的《社会科学中的研究设计:定性研究中的科学推论》(*Designing Social Inquiry: Scientific Inference in Qualitative Research*)。<sup>①</sup>对于KKV的DSI一书的讨论及批判,事实上存在两种类型的错误,这两种错误贯穿了许多方法论的讨论,并一定程度地导致了方法论研究的困惑甚至混乱,因此有必要予以指出和澄清。

第一种误区认为,定性研究和定量研究是同一种研究方法,两者之间的差异仅仅在于样本数量。这种错误是DSI导致的最大误解,该书一直暗示定量研究是比定性研究更为“先进”的方法,而定性研究的学者需要通过各种方法来增加样本数量以保证结论的有效性。KKV的这种偏见很大程度上是定量研究学者对于定性研究的误解,即认为跨案例比较是检验理论有效性的唯一方法。KKV的这类错误实质上是由于将定量方法作为主要标准,并试图将这类标准推广到定性研究的领域。第二种误区则处于另外一个极端,它认为定量研究和定性研究是两种截然不同的研究方法。定量研究和定性研究是两种不同的范式,前者通过案例间比较来进行因果推断,讨论原因的影响(effect of cause),主要实现外部有效性;而后者则通过案例内研究来实现因果,着重于讨论结果的原因(cause of effect),主要实现内部有效性。KKV第一种误区中的许多问题,已经在所谓的“后KKV时代”得到一定的解决,有大量的著作对于KKV的这类问题进行批评和纠正。<sup>②</sup>但KKV写作DSI一书的主要目的在于试图统一定性和定量两种方法的准则与逻辑,尽管KKV所采取的方法存在诸多错误,但是这种尝试本身是正确的。KKV的绝大多数批评者们忽略或者放弃了对于统一原则的追寻,过于强调两者之间的差异反而导致了方法论之间的对立与隔阂,而较少地讨论这些方法背后逻辑本身的相通之处。本文的目的在于通过探究各类方法理论背后的逻辑,讨论各类研究方法的优势和劣势的逻辑根源,借此重新审视KKV和后KKV时代对于方法论的误区和偏见,从而打破定量和定性研究之间的隔阂。

理论是科学方法的起点和终点。理论的基本逻辑是案例比较,社会科学之所以成为科学的原因在于它遵循控制与比较。布尔代数就是将人类的思维方式数学化的结果,在具体的方法层面表现为基本的逻辑思维方式,我们称之为密尔方法。密尔方法作为因果思考的基本工具,研究者因为不同的样本而选择不同的研究方法。这些研究方法主要通过对两类因果观测值的归纳来解释事物,通过控制变化来探求样本的外部和内部有效性,从而检验从原因到结果之间的因果过程。

早在上世纪70年代,阿伦·利帕特(Arend Lijphart)就已经提出了比较方法是建

<sup>①</sup> 为了便于叙述,将作者简称为KKV,将该书简称为DSI。参见Gary King, Robert Keohane and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton: Princeton University Press, 1994.

<sup>②</sup> Henry E. Brady and David Collier (eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham: Rowman & Littlefield Publishers, 2010; James Mahoney, After KKV: The New Methodology of Qualitative Research, *World Politics*, vol. 62, no. 1, 2010, pp. 120-147.

立通用的实证命题的基本方法之一（其他方法则包括实验、数据和案例分析）。<sup>①</sup> 此前国内的文献往往过度注重对方法的讨论，而关于研究设计背后逻辑及其关系的讨论则存在不足，因此本文试图以科学的研究设计为出发点，探讨理论构建过程中的案例、方法与研究设计。

具体而言，本文的框架如下：第一，讨论了研究设计的基本路径和步骤；第二，分析密尔方法在案例比较研究中的基本逻辑、特征和局限；第三，总结了单案例、小样本、中等样本和大样本研究这4种类型的研究方法，及其对应的形式逻辑和研究方法；第四，文章阐述了单一研究方法的缺陷，以及如何通过案例内比较与混合的方法来进一步提升结论的可靠性。最后给出结论，认为社会科学研究并不仅仅依靠科学方法和对操作流程的熟悉，经验判断和写作技巧也同样重要。

## 二、因果推断和密尔方法

自然科学的结论通常被认为比社会科学研究具有更少的争议和更高的可信度，其原因在于自然科学可以更好地借助于实验方法，通过控制变量来不断重复结果。而社会科学由于条件的限制或者具体历史的研究，多数情况下无法进行大规模的重复实验，因此社会科学的核心思想是通过控制变量来达到准实验状态。这些都要通过案例比较来实现。比较是控制变量和建立变量间普遍性关系的方法，它是一种归纳推理。约翰·斯图亚特·密尔（John Stuart Mill）最早在他的《逻辑体系》一书中提到了5种归纳推理的基本方法，被称之为“密尔五法”：求异法、求同法、求同求异法、共变法和求余法。<sup>②</sup>

（一）求异法，即通过最小化差异来求异，在密尔的体系中，所有方法都归于求异法。<sup>③</sup> 由于社会科学经常无法人为地控制变量之间的差异，因此就只能通过相似案例的比较来接近实验状态。求异法的核心逻辑就是，在其他条件一致时，导致结果不同的差异即为原因。求异法成立的前提是严格的准实验状态，否则就可能因为无法消除竞争性解释而降低理论的可信度。<sup>④</sup>

案例1:  $A+B+C+D+E \rightarrow Y$

案例2:  $\sim A+B+C+D+E \rightarrow \sim Y$

<sup>①</sup> Arend Lijphart, Comparative Politics and the Comparative Method, *American Political Science Review*, vol. 65, no. 3, 1971, p. 682.

<sup>②</sup> John Mill, *System of Logic: Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, New York: Harper & Brothers Publishers, 1898, pp. 478-537.

<sup>③</sup> [瑞士]丹尼尔·卡拉曼尼：《基于布尔代数的比较法导论》，蒋勤译，格致出版社，2012年，第13页。

<sup>④</sup> Alexander George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences*, Massachusetts: MIT Press, 2005, p. 165.

在上述两个案例中,如果因素 A 的差异导致了结果 Y 的差异,就可以认为 A 是 Y 的原因。求异法的优势在于它可以依赖较少的案例,理论上来说只需要  $2^n$  个案例就可以通过 n 个变量来解释不同的结果。求异法同样存在逻辑和实际操作的问题。首先,求异法对于准实验状态存在高度的依赖,由于社会科学中几乎不可能找到两个只有条件 A 不同而其他因素都一样的案例或样本,通常只能对于变量进行近似的控制,因此求异法往往会存在较多的竞争性解释,即被作者所忽视的差异会削弱解释变量的因果解释。因此,研究者在最大程度地控制差异之后,往往还需要解释为什么其他的差异(竞争性解释)并不重要。除此之外,仅仅停留在联列表的求异法,在逻辑上仍然是一种相关性的推断,还需要通过案例研究来增强其因果解释。

(二)求同法,即通过最大化差异来求同,其核心逻辑是在其他条件都具有极大差异的情况下,如果仍然能够达到相同的结果,事物间的共性即为其原因。

案例1:  $A+B+C+D+E \rightarrow Y$

案例2:  $A+F+G+H+I \rightarrow Y$

在以上的两个案例中,在其他原因不同而导致相同的结果时,可以认为原因 A 导致了结果 Y。求同法的优势在于其能够在存在较大差异的时候仍然维持较少的变量,但求同法是密尔方法中逻辑较弱的一种方法,其缺陷在于,如果无法说明其他因素不重要,由于无法解决变量和结果之间存在的“多重因果性”,即各类变量/原因相加导致的相同结果,就会降低理论的解释力。<sup>①</sup>比如毕业生的薪水是由专业、能力和学历等因素综合决定的,显然如果依据单一要素进行求同往往得出的结论可能是错误的。事实上很多传统的谚语都表达了上述观点,即不同因素的组合可以达到相同,例如“八仙过海,各显神通”“条条道路通罗马”,等等。因此通常情况下,在定性的研究中,如果仅仅停留在对于少量因素的求同,就无法得出正确的结论。

(三)由于求同法的逻辑较弱,密尔提出了求同法和求异法的共同使用(Joint Method of Agreement and Difference),或者称之为间接求异法(Indirect Method of Difference)。<sup>②</sup>即在单一小组内部求同,而在小组之间求差异。假设案例用一个变量的差异分为两个小组:在其中一组中,因素 A 出现(记作 A),而在另一组中,因素 A 不出现(记作  $\sim A$ )。从而能够在这两个小组之间使用求异法,即只有 A 的差异而其他因素相同,最终导致了不同的结果;与此同时,小组内部虽然其他因素存在较大的差异,但是因为共同原因 A 又

<sup>①</sup> Alexander George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences*, pp. 156-157; Charles Ragin, *Fuzzy-Set Social Science*, Chicago: University of Chicago Press, 2000; Charles Ragin, *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, Chicago: University of Chicago Press, 2008.

<sup>②</sup> John Mill, *System of Logic: Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, New York: Harper & Brothers Publishers, 1898, pp. 488-489.

达成了相同的结果（即  $A \rightarrow Y$  且  $\sim A \rightarrow \sim Y$ ），这就具有较强的说服力可以认为 A 是结果 Y 的原因。

（四）共变法是通过控制其他条件不变，观察 A 和 Y 的共同变化，可以认为两者之间存在因果性或者由于共同原因导致了变化的一致性。共变法在大样本的研究中具有较强的可信度，但是它的缺陷是仅仅表达了一种统计上的相关性。在自然科学中，研究者可以通过控制和重复试验，明确观察到 A 和 Y 之间的因果性。但是在社会科学中，研究者往往只能事后通过既有数据进行判断，因此只能观测到相关性。共变法往往无法直接区分因果倒置和内生性问题，即无法判断自变量和因变量究竟哪个是原因，哪个是结果，或者互为因果，往往社会科学中存在较多的互为因果的案例，例如政治与经济。同时，共变法也无法区分伪相关，因为自变量和因变量之间的相关性可能是由于共同原因导致的，而两者之间不存在因果关系。<sup>①</sup> 例如游泳的人数和买雪糕的人数之间存在相关性，但是两者是因为天气炎热导致的，没有因果联系。

（五）求余法，即认为存在原因  $A+B+C$  导致了结果  $a+b+c$ ，如果可以确信 A 是 a 的原因，而 B 是 b 的原因，则可以推断 C 是 c 的原因。上述 4 种方法属于归纳法，而求余法则会在演绎逻辑，需要通过推理得出结论。求余法在个案的研究中存在较大的优势，可以通过排除法迅速找到现象的原因，正如福尔摩斯所说的那样：“排除一切不可能的原因，剩下的再令人难以置信，那也是真相。”求余法同样存在一定的局限，它需要基于原因之间互相独立的假设，即原因 A、B、C 之间不存在互相干扰，否则就无法通过排除法进行推理。

密尔五法在本质上都属于排除法，属于一种形式逻辑，即通过排除其他干扰因素之后，最终推导出自变量和因变量之间的关系。而形式逻辑的数学基础是布尔代数，在数学化之后，就是通过“或”（符号表示为 +）、“且”（符号表示为 \*）、“非”（符号表示为  $\sim$ ）的运算来寻求因果解释。

### 三、案例比较与研究方法

前文阐述了案例研究的基本逻辑，在此基础上文章进一步讨论案例比较的方法。对于案例选取的规则恰恰是基于对第二部分所述的各类研究方法的优势和缺陷。在讨论案例之间的差异时，需要明确案例选取的时空边界。一方面，这是由求异法的逻辑特征所决定的，求异法是基于最小化差异来进行因果推断，故而会面临诸多的竞争性解释。因此在案例的时空背

---

<sup>①</sup> Daniel Steel, *Across the Boundaries: Extrapolation in Biology and Social Science*, Oxford: Oxford University Press, 2008, p. 177.

景差异过大时,时空差异本身就会构成对理论解释的威胁,像加州学派对大分流的讨论,事实上就忽略了中国和欧洲(或者英国)截然不同的背景,这些差异才可能是导致结果不同的原因。<sup>①</sup>另一方面,即便是同一概念,在不同的时空情境中也可能赋予其不同的涵义,例如大航海时代的海权和当代的海权。<sup>②</sup>时空情境的限制往往也缩小了案例选择的范围,比如在某一个时间段的特定地区内选取案例进行比较。

因此,在案例选择中需要避免几方面的错误。第一,避免简单的举例法,比如许多中学里的语文议论文往往就是从历史长河中选取几个例子来验证观点。事实上,多数情况下我们总能够找到符合理论的例子,当然这种“理论”也往往会存在诸多反例。第二,避免基于“情怀”或者便利的案例选择,例如,根据作者自身的语言基础而非方法论的需要来进行案例选择。当然,文章选题本身基于情怀或者便利,这并没有问题,但是案例选择不能因为对特定国家特别感兴趣或者熟悉来进行,而是要在对总体的样本和理论把握的前提下,依据最小化差异的原则进行筛选。否则,极有可能会出现无法满足最大相似性的情况,从而损害了案例比较的因果解释力。第三,避免只关注重大事件,没有发生的事件——往往是反面案例——同样重要。例如在研究革命时,西达·斯考切波(Theda Skocpol)就不仅仅关注法国大革命这样的重大事件,同样关心普鲁士为什么没有发生革命。<sup>③</sup>

在确立了研究的基本问题之后,恰当的案例选取是实证研究的核心环节,而案例的选择本身也决定了采取何种研究方法,需要以问题而非研究方法为导向。问题性质和时空范围的限制往往已经决定了可以选择的样本数量,而样本数量又进一步限制了研究方法的选取。詹姆斯·马奥尼(James Mahoney)和加里·葛尔兹(Gary Goertz)提出了案例选择的两个条件范围(Scope Condition)和可能性原则(Possibility Principle)。<sup>④</sup>条件范围的假设是求异法逻辑成立的关键,即案例之间要具备较高的相似度,即符合“同质性假设”。案例选择的“可能性原则”讲述的是要选取有可能成功的案例作为负面案例,避免将完全不可能发生的案例误认为负面案例,例如在讨论贸易是否有助于促进平时,显然用两个完全不可能发生战争的国家(例如秘鲁和蒙古)作为负面案例。假设某个理论要用变量A和变量B来解释结果Y,那么A和B都出现(A=1且B=1)就是结果Y出现(Y=1)的充分必要条件,那么在现实情况下,作为观测值的样本存在一共8种情况,即变量和结果分别为0或1的情况。

① 参见赵鼎新:《加州学派与工业资本主义的兴起》,《学术月刊》2014年第7期。

② 这部分具体的讨论参见Tulia G. Falleti and Julia F. Lynch, Context and Causal Mechanisms in Political Analysis, *Comparative Political Studies*, vol. 42, no. 9, 2009, pp. 1145-1151; 叶成城、黄振乾、唐世平:《社会科学中的时空和案例选择》,《经济社会体制比较》2018年第3期。

③ [美]西达·斯考切波:《国家与社会革命》,何俊志等译,上海人民出版社,2013年。

④ James Mahoney and Gary Goertz, The Possibility Principle: Choosing Negative Cases in Comparative Research, *American Political Science Review*, vol. 98, no. 4, 2004, pp. 653-669.

表1. 双变量情况下的案例类型

	Y = 0		Y = 1	
	B = 0	B = 1	B = 0	B = 1
A = 0	无关案例	负面案例	反例	反例
A = 1	负面案例	反例	反例	正面案例

资料来源：笔者自制

如上表所示，观测值根据对理论的支持情况分为4种类型，正面案例和负面案例支持文章的理论，两种共同描述了自变量组合对于结果的充分必要条件，即只有A和B共同出现时结果才会出现，两者缺一不可。两者都缺失的案例为无关案例，它既不能证伪理论，对于理论的支持也非常微弱。而出现与上述4种类型不同的理论预期时，例如当A = 0且B = 0时，如果结果Y出现了，就会形成对理论的重要威胁，往往意味着有可能遗漏了重要的变量。

在确定了样本类型与结果的关系之后，根据样本数量的多少，研究类型分为单案例研究、小样本研究、中等样本研究和大样本研究。首先，单案例研究往往因为案例的特殊性较强，缺乏相似的案例进行比较，因此只能对个案研究进行深度挖掘，通常采用过程追踪的方法。案例研究的作用是发现理论，其优势是可以具备更加细致的过程追踪，但是单案例研究所发掘的理论的推广性经常是有限的。这类研究通常做得更为精细，但也更加具有不可证伪性，容易提供一种全要素的解释，但随着样本量尤其是反例的增加，就越发容易出现概率性的解释。<sup>①</sup>其次，小样本研究是定性研究中最常见的情况，例如在限定时空范围后，可能会出现小样本，但却是“全样本”，即符合条件的样本只有2~12个左右。小样本研究更加注重案例本身的细节，以“深度”见长。小样本研究通常采用定性分析的方法，它更容易发现和提出理论假说，也可以借助控制变量法来检验理论。小样本的案例中如果出现“准实验状态”，那么定性分析就非常有效。<sup>②</sup>第三，中等样本的研究是查尔斯·拉金（Charles Ragin）所提出的，他认为用定性比较分析的方法可以解决案例研究和回归分析无法解决的中等样本，即样本数量在12~30之间的情况。<sup>③</sup>在样本数量多到无法使用小样本的比较分析，样本数量又不足以进行回归分析时（例如样本数量少于30个），采用定性比较分析（Qualitative Comparative Analysis, QCA）就更为有效。中等样本研究更加接近于定量方法而非定性，

<sup>①</sup> John Gerring, *Case Study Research: Principles and Practices*, Cambridge: Cambridge University Press, 2007, pp. 193-195.

<sup>②</sup> John Gerring, *Case Study Research: Principles and Practices*, Cambridge: Cambridge University Press, 2007, p. 57.

<sup>③</sup> Charles Ragin, *Fuzzy-set Social Science*, Chicago: University of Chicago Press, 2000; Charles Ragin, *Redesigning Social inquiry: Fuzzy Sets and Beyond*, Chicago: University of Chicago Press, 2008.

因为它更加重视样本的同质性和外部有效性，往往只能够检验假设，通常也很难产生机制性的解释。第四，大样本研究通常需要样本数量大于 30 以上，以回归分析为主要研究方法。大样本研究更加具备广度，通常只能对设定好的有限假设进行检验，但是具备较高的置信度。<sup>①</sup>大样本研究建立在同质性假设的基础上，它更加重视样本的外部效度和平均效应，它更多地考察样本之间的共变，因此它只能够进行因果推断而非因果解释，它的因果效应仅仅是建立在相关性的统计描述之上的，因而其因果性更弱。

表2. 案例数量和研究特性

案例类型	单案例	小样本	中等样本	大样本
1. 样本量	1	2~12	12~30	≥30
2. 假设	产生假设	产生和检验假设	检验假设	检验假设
3. 有效性	内部	内部为主	内部+外部	外部
4. 因果视角	因果机制	因果机制	部分因果机制	因果推断
5. 范围属性	深度	深度	深度+广度	广度
6. 总体案例	—	差异性	同质性	同质性
7. 因果解释	—	强	中	弱
8. 研究方法	过程追踪	质性比较	定性比较分析	定量分析
9. 形式逻辑	求余法	求余法/求同求异法	求同求异法/共变法	共变法

资料来源：笔者自制

如表 2 所示，不同的样本量会涉及不同的研究方法，单案例、小样本、中等样本和大样本研究分别对应了过程追踪、质性比较、QCA 和定量分析 4 种方法。

### （一）单案例与过程追踪

过程追踪是对案例本身进行讨论和研究，加里金认为这是在增加案例内部的观测值。马奥尼提出了过程追踪中的几个基本步骤：包括了三种检验方式，分别是寻找线索、进行必要性检验和进行充分性检验。上述三种检验方式属于因果性检验，马奥尼认为过程追踪中还需要进行（关于事件的）存在性检验。<sup>②</sup>因果性检验即是常见的因果分析，更多地是社会科

<sup>①</sup> John Gerring, *Case Study Research: Principles and Practices*, Cambridge: Cambridge University Press, 2007, p. 41.

<sup>②</sup> James Mahoney, The Logic of Process Tracing Tests in the Social Sciences, *Sociological Methods & Research*, vol.41, no.4, 2012, pp.574-583.



学家的任务；而存在性检验则是对特定争议事件的甄别，经常是历史学家的任务。相比起历史学，政治学和经济学的研究会更多地使用二手资料，更加关注不同因素的组合如何导致了某种行为或结果。方法论的研究者往往用侦探小说中的破案过程来类比案例研究中的过程追踪，即研究者如同侦探破案一般地阐释自变量通过何种机制来导致因变量。大卫·科利尔（David Collier）用了柯南道尔的《福尔摩斯探案集》的《回忆录》中的《银色马》来讲述如何进行过程追踪。<sup>①</sup>他总结了过程追踪的4个基本步骤：首先，通过可疑线索提供研究基础；其次，通过必要性检验来排除不可能的原因；再次，寻找充分条件确定假设；最后，综合此前的线索推理得出充分必要条件。

## （二）质性比较往往用于小样本研究

小样本的定性分析的步骤大致如下：1. 声明案例选择的理由（排除无关案例）；2. 列出变量的类型交叉表格，在定性研究中，受到样本的限制，核心变量数目往往不能过多；3. 区分案例的作用，对正面案例和负面案例进行对比，检验结果是否在正面案例中出现，而在负面案例中不出现；4. 解释与理论预期不相符合的反例；5. 通过过程追踪来明晰从自变量到因变量的因果机制。假设在双变量的情况下，研究者往往就用两个自变量 A 和 B 形成一个  $2 \times 2$  的表格来研究自变量与因变量的关系。例如丹尼尔·兹伯拉（Daniel Ziblatt）的《构建国家》一书中，用地区性制度效率（自变量 A）和联邦主义效率（自变量 B）两个因素去构建其理论模型，探讨能否形成，他发现只有具备较高地区制度效率（A=1）和联邦主义的国家（B=1），才能够构建联邦主义国家（Y=1）。<sup>②</sup>这种情况下，通常可以认为 A=1 且 B=1 是 Y=1 的充分必要条件。

	A=0	A=1
B=0	Y=0	Y=0
B=1	Y=0	Y=1

由于观测值较少，忽略案例内部特性的讨论往往缺乏足够证据，它需要以案例内部的深度挖掘为主，通过过程追踪和质性比较可以得出较强的因果解释。例如  $X \rightarrow Y$  只有 1~3 个样本，可信度不高；但是如果在这几个案例中，都验证了特定机制，比如在  $X1 \rightarrow X2 \rightarrow X3 \rightarrow Y$  形成证据链，并在因果链上都能找到证据，解释的有效性就大大增

<sup>①</sup> David Collier, *Understanding Process Tracing*, *PS: Political Science & Politics*, vol. 44, no. 4, 2011, pp. 823-830.

<sup>②</sup> Daniel Ziblatt, *Structuring the State: The Formation of Italy and Germany and the Puzzle of Federalism*, Princeton: Princeton University Press, 2006.

强了。

### （三）定性比较分析

拉金早在 1987 年将布尔代数和集合理论结合起来，发展出二分变量的 QCA 技术。<sup>①</sup> 定性比较分析结合定性和定量分析两种方法的长处，将研究的各个案例都视作整体，并分析案例中的解释条件及其组合，从而有效处理蕴涵在案例中的复杂因果关系。<sup>②</sup> QCA 在政治学和历史社会学领域使用得更为广泛，因为这些领域的样本数量通常较少，而在经济学领域通常样本较多，因此 QCA 的应用相对较少。定性比较分析的主要方法有清晰集分析、模糊集分析和多值分析，主要处理自变量和因变量全部都是虚拟变量的数据，而模糊集理论的运用则一定程度使其可以处理连续变量。<sup>③</sup>

定性比较分析更重视案例间的平均效应，探讨的是概率而非覆盖率的问题。在中等样本并且存在较多解释变量时，定性比较分析能够提供有效的分析。以多值分析为例，唐睿和唐世平讨论了冷战后东欧国家民主转型的例子，自变量为是否为伊斯兰国家、是否为前苏联加盟共和国、是否有民主经历、是否有独立国家经历、经济发展水平是否较高、资源配置扭曲程度是否较高。<sup>④</sup> QCA 的步骤也是一个寻找近似充要条件的过程，但它更加侧重于平均效应。首先，QCA 需要对单个变量进行必要性检验，如果非 A 是非 Y 的充分条件（吻合度大于 0.9），即假如  $A = 0$ ，结果必然基本上不出现；其次，剔除作为必要条件的变量和必要条件等于 0 的样本重新检验，即剔除伊斯兰国家和二战前的苏联加盟国再度进行检验，察看结果出现（ $Y = 1$ ）可能性最高的组合，寻找此前提下的充分条件；最后，结合充分和必要条件，通过布尔代数合并充分条件，计算出最后的充分必要条件。

### （四）定量分析

定量分析是经济学和社会学中最为常用的方法，它的基本逻辑是基于共变法和对频数的统计。在实证研究中，定量分析通常是基于推断性统计，回归分析则是最为常见的定量方法，定量方法大致分为如下几个步骤。首先，确定研究问题的自变量和因变量，形成待检验的假设。定量研究对于数据和问题的清晰程度要求比上述其他三种研究更为严格，而且对于因果关系的清晰程度的要求也更高，除了难以解决范式类的研究之外，通常也很难解决特别复杂的因果关系。<sup>⑤</sup> 其次，测量变量和生成数据之后，对数据进行处理和描述性统计，例如给出变量之间的相关系数、方差、极值等数据。如果文章的理论正确，并且具有较高质量的数据，

<sup>①</sup> Charles Ragin, *The Comparative Method. Moving Beyond Qualitative and Quantitative Methods*, Berkeley: University of California, 1987; Charles Ragin, *The Logic of the Comparative Method and the Algebra of Logic*, *Journal of Quantitative Anthropology*, vol. 1, no. 4, 1989, pp. 373-398.

<sup>②</sup> 迟永：《美国介入领土争端的行为——基于模糊集定性比较分析的解释》，《世界经济与政治》2014 年第 10 期。

<sup>③</sup> 刘丰：《定性比较分析与国际关系研究》，《世界经济与政治》2015 年第 1 期。

<sup>④</sup> 唐睿、唐世平：《历史遗产与原苏东国家的民主转型——基于 26 个国家的模糊集与多值 QCA 的双重检测》，《世界经济与政治》2013 年第 2 期。

<sup>⑤</sup> 庞珣：《国际关系研究的定量方法：定义、规则与操作》，《世界经济与政治》2014 年第 1 期。

通过描述性统计，例如趋势图可以在一定程度上作出初步的判断。再次，建立统计模型进行参数估计和假设检验。以 t 检验为例，在控制了基本变量后，如果自变量的系数为 0 的概率低于一定程度（如 0.05 或者 0.01），则可以认为能够拒绝原假设，从而证实其结论。詹姆斯·雷（James Ray）提出回归分析的 5 个基本注意事项，包括不能控制中介变量，区分补充性解释和竞争性解释，不能过度增加控制变量，不要控制和解释变量在定义上非常接近的因素，控制可能的时间和空间差异。<sup>①</sup> 又次，对检查模型结果进行稳健性检验，通过变换测量方式和控制变量察看模型是否稳健。最后，对统计结果进行分析，通常情况下，由于定量分析往往只能解释原因和结果之间的相关性，对于结果的分析往往依赖于理论和专业知识，而对过程的推测本身仍然存在不确定性。

#### 四、研究方法的特征与共同使用

需要看到的是，上述 4 种研究方法的逻辑是不一致的。个案追踪和小样本比较都可以视为定性研究，而 QCA 和回归分析则更接近于定量研究。两种最核心的差异在于寻找因果解释到底是试图寻找“结果的原因”（Causes of Effects），还是仅仅是试图甄别“原因的影响”（Effect of Causes）。<sup>②</sup> 定量研究依赖于数据集的观测值（Data-Set Observation, DSOs），而定性研究则更加依赖于因果过程的观测值（Causal-Process Observation, CPOs）。<sup>③</sup> 假设理论预期出现 10 个 CPOs，在实证研究中出现 5~8 个线索，就基本可以判断这种理论的可信度较高。事实上，5~8 个线索已经足够让假说成立，在 CPOs 的研究中，80% 的置信区间已经很高，追加额外的证据边际贡献已经很少，无需像经济学家一样追求 95% 或者 99%。这种逻辑的差异也导致了定性研究和定量研究在样本选择上的差异。后者通过随机性使定量分析的残差符合正态分布，因此强调选择的随机性和无偏性，例如不能够根据因变量来进行样本选择。<sup>④</sup> 而前者则是基于对因果机制的探索，事实上很多基于对“半负面案例”的选择不可避免地会带来样本选择偏差。<sup>⑤</sup> 在这些方法中，过程追踪和质性分析更加偏向于案例内的效度，而它们的差异则在于，质性分析一定程度依赖案例间的比较通过求同 / 求异法来得出

---

① James Lee Ray, Explaining Interstate conflict and War: What Should Be Controlled For? *Conflict Management and Peace Science*, vol. 20, no. 2, 2003, pp. 1-31.

② Gary Goertz and James Mahoney, *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton: Princeton University Press, 2012, p. 43; 唐世平:《超越定性与定量之争》,《公共行政评论》2015 年第 4 期。

③ James Mahoney, After KKV: The New Methodology of Qualitative Research, *World Politics*, vol. 62, no. 1, 2010, pp. 120-147.

④ Barbara Geddes, How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics, *Political Analysis*, vol. 2, 1990, pp. 131-150.

⑤ 这部分的具体讨论参见周亦奇、唐世平:《“半负面案例比较”:北约与华约的命运为何不同?》,《世界经济与政治》2018 年第 12 期;叶成城、唐世平:《基于因果机制的案例选择》,《世界经济与政治》2019 年第 10 期。

因果推断，但是离开了对案例的深入研究来谈平均效应，质性分析会出现“小样本谬误”。而 QCA 和定量分析，则更加重视案例间的平均效应，即通过统计自变量的频数来讨论它们对于因变量的平均影响，单个案例起到的作用微乎其微。而 QCA 和定量分析的差异在于对频数的统计，QCA 重在寻找类型组合导致结果出现的概率；而定量方法如回归分析则建立在高斯分布的基础上，通过拒绝自变量或变量组合系数为 0 的原假设来确立自变量对因变量的影响。

上述的 4 种样本类型对应了 4 种基本方法，但是这种对应和界线并不是绝对的。而这 4 种方法各自都存在一定的缺陷，这种缺陷源自于它们对应的密尔方法，采用混合方法可以在一定程度上弥补各自的缺陷并增强解释力。

混合方法则是通过结合或者部分结合定量和定性方法来避免单一方法的缺陷。<sup>①</sup>其主要目的包括寻求不同方法的研究结果的聚合、扩大特定项目的广度和范围、互相补充或者发现悖论、通过不同方法推进研究，等等。<sup>②</sup>约翰·克雷斯威尔（John Creswell）基于不同的研究特征总结了 6 种基本混合方法的策略方法，通过不同方法之间的互相印证提高理论的可信度。<sup>③</sup>需要看到的是，对于不同混合策略的选择，往往也受到样本特性的影响，例如在样本较大时，定量研究的优势会更加明显，反之亦然。假设研究的自变量 X 通过中介变量 M 导致结果 Y，即  $X \rightarrow M \rightarrow Y$ 。定量为主的研究往往通过大样本数据证明 X 和 Y 之间的共变关系，由于数据分析无法阐述案例的因果机制，因此需要用定性方法去补充，展示  $X \rightarrow M$  和  $M \rightarrow Y$  的具体过程。同理，在定性为主的样本研究中，由于数量过少，就无法成立案例之间的有效性，但是仍然有可能通过案例内比较来证明  $X \rightarrow M$ ，假如能够有足够多的定量数据来证明，X 和 M 之间的相关性，那么就会成为整个定性研究中非常重要的证据。通常情况下，只有在样本具备一定规模但仍然不是很多的情况下（例如数百个），且拥有足够的资源可以对每一个样本进行案例研究，否则很难严格实行两种方法的混合应用。

## 五、结论

前面介绍了社会科学中的研究设计和案例比较的基本逻辑和方法，无论是定量还是定性方法，它们都有着各自的优点和局限，在使用任何一种方法或者混合方法时，研究者都要明白它们的适用范围，而样本数量和样本特性则是对适用范围最基本的限制。社会科学中的研

---

<sup>①</sup> Evan S. Lieberman, Nested Analysis as a Mixed-Method Strategy for Comparative Research, *American Political Science Review*, vol. 99, no. 3, 2005, pp. 435-452.

<sup>②</sup> [美] 阿巴斯·塔沙克里、查尔斯·特德莱：《混合方法论：定性方法和定量方法的结合》，唐海华译，重庆大学出版社，2010 年，第 41 页。

<sup>③</sup> [美] 约翰·克雷斯威尔：《研究设计与写作指导：定性、定量与混合研究的路径》，崔延强译，重庆大学出版社，2007 年，第 168-173 页。

究设计,既是一门科学,也是一种艺术。一方面,社会科学不是诠释,不论是定性还是定量,都要有严格的内部逻辑过程,具有一定的操作程序和流程,通过数据集和因果过程的观测值作为实证证据来增加因果分析的可信度,即尽可能地提高 X 是 Y 的原因的可能性。社会科学作为一门科学,具有一定的可重复性,即如果采用特定的测量、比较和因果分析的流程,无论是基于 DSOs 还是基于 CPOs 的研究,只要研究的史料和数据没有发生根本性的变化(如特定事件被证明被严重误解或测量方法存在偏差),不同人进行定量或者定性研究得出的结论是相似的。当然,往往对基于 DSOs 的研究进行重复时,结果会更加相近,因为不同的研究者对于特定的 CPOs 的理解更有可能产生分歧。

另一方面,社会科学又是一门艺术,科学方法也仅仅是诸多研究方法中的一种。研究者对于研究问题和解释变量的选择,往往受到自身的经历、价值观念或者语言文化背景的深刻影响。同时,样本数量并不完全局限于方法的选择,即便是单一案例,也可能在微观层面(例如次国家层面)存在大样本分析的必要;而随着 QCA 软件的不断改善,同样可以处理诸如 100~200 个样本的研究,此时就需要权衡采用 QCA 还是定量分析,或者两者兼用。除此之外,研究论文本身也有篇幅的限制,在有限篇幅内往往无法穷尽方法和案例,因此如何进行谋篇布局也需要经验和艺术。

(责任编辑:俞平)

## Causal Explanation In Social Sciences: Trade-off Among Logic, Samples and Methods

YE Chengcheng

**Abstract:** With the development of empirical research in the field of social science, comparative study has become one of the most important research methods in social sciences. Rigor empirical research requires scientific research design and follows procedural steps for causal explanation, including the selection of independent variables and dependent variables, theory construction, and hypothesis test. Causal explanation needs to explore the formal logic named Mill's Methods which are intended to illuminate issues of causation. Mill's five Methods thus correspond to four basic case study methods: process tracing, case comparison, qualitative comparative analysis (QCA), and quantitative research. The first two methods focus on the observations within the case, while the latter two methods focus on the average effect between cases. They have their own connections, advantages and disadvantages in different types and numbers of cases, therefore, researchers can explore mixed methods to avoid the defects of a single method.

**Key Words:** social sciences; Mill's Methods; case comparison; causal explanation; trade-off; quantitative & qualitative study