

# 龙小宁、宋健：如何在经济分析中进行因果识别（上）

## ——关于损害赔偿计算的对谈之 4

对谈人：

龙小宁 厦门大学“长江学者”特聘教授、本刊专栏作者

宋 健 江苏省高级人民法院原资深法官、全国审判业务专家、本刊专栏作者

宋 健：龙老师，在前三次对谈中，我们讨论了有关损害赔偿计算方式的一些基本问题。对谈 1 重点讨论了精确推导的数学计算对于损害赔偿计算的局限性，并进一步讨论了引入统计学计算的重要性；对谈 2 重点讨论了引入经济学计算的重要性；对谈 3 重点探讨了如何理解经济分析的科学性。以上问题的探讨，有助于解释为什么损害赔偿计算以及其他类型案件中引入经济分析及经济学计算的重要性及必要性，例如涉及 SEP 许可费率的计算、反垄断相关市场的分析等。

在对谈 3 中，最后留待讨论的问题是，如何进行理论模型中的“变量”选择和调整？您当时提出，这个问题实际涉及到因果识别的重要性的和社会科学的“实证革命”，希望听到你的进一步分析意见。

龙小宁：在上一次对谈中，我们讨论到人文学者对经济学等社会科学分析的批判，很多人可能认同下面的固有观念：因为关于人的行为和人类社会的相关研究具有人文学科的特征，所以没有办法用科学的理论研究方法去进行分析；尤其是，经济学理论建模过程中引入了这么多抽象因素，当然不能完整真实地反映人类社会中的决策过程，所以这一定是不正确的。

那么如何回应这样的批评意见呢？一个回应的角度是，基于上一次对谈中关于科学理论的一般性讨论，任何理论都不可避免会涉及一定程度的抽象；但另一个更有益的回应的角度是，可以引入科学研究中的实证分析环节，即针对被质疑的理论假设，通过分析现实中的数据规律来检验它是否能够准确地反映人类的决策过程。

大部分经济学家进行的研究工作都可以归入这一类实证研究：他们可以针对不同理论进行实证检验，挑选出其中更符合实证规律的模型；除了对理论假设进行定性检验之外，他们还可以定量测度理论中涉及的影响效果的大小。除了以上研究工作外，实证领域的经济学家还在另外一个非常重要的领域做出了突出贡献，这就是因果关系的识别！

**宋 健：**说到因果关系，法律人的确特别关注因果关系的分析。有观点认为，知识产权损害赔偿难以计算，原因正在于因果关系较难确定，即在个案中怎么就能确定权利人的损失是因为侵权行为造成的，而不是因行情下跌等市场因素导致的，抑或两方面因素都存在，对此我们应当如何加以识别、区分及测算呢？这是否就是经济学研究中的因果关系识别问题呢？

**龙小宁：**对的，这同样也是经济学关注的问题。简单来说，经济分析中的因果关系识别就是指在分析中把影响某个经济结果的最重要决定因素提取出来，作为研究的关注焦点。举例来说，在研究人力资本生产过程时，如果用收入来衡量人类资本水平，它的最重要决定因素是受教育的时长，相应的经济分析中就需要进行受教育时长与收入之间的因果识别。

就像宋老师上面提到的，需要解决的问题是，如何对待其他那些没有考虑到的决定因素？这个问题同时涉及两个方面，一个是理论方面，另一个是实证方面，也即与数据有关的方面。从理论角度讲，我们需要在理论模型中解释为什么这些被忽略的其他因素不会系统性地改变理论所关注的主体规律，比如受教育时长与收入之间的关系；从实证角度看，我们的处理方法是尽可能全面收集这些其他因素对应的衡量指标，并在实证分析中考虑这些因素的影响，比如在多元线性回归模型中加入这些指标作为控制变量。通过在实证分析中“控制”或者“剔除”这些其他因素的作用，我们便可以识别出受教育时长与收入之间的因果关系，也即理论模型中所关注的主体规律。

具体到实证分析中，我们需要得到一个模型，来给出各种影响因素如何结合

在一起决定收入水平。这个模型可以用一个等式  $y=a+bx+cz$  来表示，等式中左侧  $y$  是需要解释的变量，称作被解释变量，是我们关心的结果，例如年收入；右手侧是解释变量，又包括两类变量，第一类变量  $x$  是理论模型中主要的决定因素，例如受教育时长，它们的选取决定于模型的研究目的，也源于模型简化的需要；<sup>1</sup>第二类变量  $z$  是可观测的控制变量，它们因为不是经济理论模型研究的重点，所以没有作为主要变量包括在模型中，但既然可以观察到，那么就可以在实证模型中加以控制来帮助改进实证估计的效果。<sup>2</sup>

除上述两类变量外，还有第三类变量，也是让我们最为纠结的，就是与今年诺贝尔经济学奖授予的研究成果有关的一类变量。这类变量是在理论模型中会对被解释变量起到关键影响作用，但却无法在实证研究中直接控制的变量，因为在现实中无法直接观察到它们对应的衡量指标，它们被称为“混杂因素”，也是因果关系识别中的核心挑战。

仍以上述教育水平对收入的影响研究为例。理论模型推导出教育水平对收入有正向的影响，其中收入是被解释变量，教育水平是解释变量，而可观测的控制变量包括年龄、性别等。此外，还有各种混杂因素，包括个人的天分、自律程度等等。上述可观测的控制变量和混杂因素都可能会对收入水平产生影响，但混杂因素除了不可直接观测之外，又与教育水平呈正相关关系，因为天分、自律等因素恰恰跟教育水平有明显的正相关关系。如果忽略这些混杂因素，只看教育水平的影响，然后得出收入高是因为教育水平引起的结论，就会把所有贡献都归结到教育水平上，而忽略其他混杂因素变量同时也在发生作用，此时因果判断就出现了偏差。

这种现象在经济学中称为“共时性”，即指两个变量的值可能是由其他一些变量同时决定的，因此不能推导这两个变量之间具有因果关系。在我们的例子中，个体的天分、自律等同时决定教育水平和收入水平这两种结果，因而我们不能准

---

<sup>1</sup> 关于理论模型为何需要简化，参见对谈 3。

<sup>2</sup> 因为模型是基于观察数据得到的，数据和对应的模型都具有一定的随机性，所以我们把得到模型的过程称为“估计”。

确推断出教育水平对收入水平具有决定作用。

“共时性”是“内生性”的一种，“内生性”的意思是说我们视为原因的变量本身也是被其他因素决定的结果，所以是模型中“内生”的，而不是单纯“外生”的原因，比如上面模型中作为收入这个结果的决定因素（或说原因）的教育水平变量本身同时还受其他因素的影响而决定。内生性有两种表现形式，一种是上面讨论的“共时性”，另一种表现形式则是反向因果，也即解释变量反而可能是结果，而被解释变量则是原因，例如在上面的讨论中，收入水平的提高可能令成人教育时长增加，受教育时长变成了受收入影响的结果。但涉及混杂因素时，需要解决的内生性问题是“共时性”。

那么，如何在不能直接观察到混杂变量取值的情况下，将它们的影响纳入实证分析中，进而恰当地排除它们的影响，从而正确地进行因果推断呢？获得今年诺贝尔经济学奖的几位学者，正是在这个研究领域做出了卓越的贡献。

**宋 健：** 您能具体讨论一种或几种经济学中用于进行因果推断的方法吗？

**龙小宁：** 好的。在几位诺贝尔经济学奖得主开发、应用和推广的方法中，我着重讨论两种，也是在其他法域诉讼案件中已经有过应用实例的方法，一个是“工具变量法”，另一个是“双重差分法”。<sup>3</sup>

关于工具变量法的思路是这样：比如在上述例子中，考虑到学生的个体天分会影响收入，而且天分又跟教育有正相关性，如果只关注教育与收入的关系，可能会错误地将天分对收入的贡献归功于教育。为了解决这一问题，可以做这样一个实验设计，即让个体接受教育的时间长度跟天分、自律程度等因素都没有关系，这就是所谓的“随机实验”的方法。具体来讲，有一群孩子，天分、自律程度等存在高低差异，需要随机决定他们受教育时间的长短。需要注意的是，受教育时

---

<sup>3</sup> 工具变量法是欧盟委员会 (European Commission) 2003 年 Lagard`ere - VUP 调查和英国竞争委员会 2007 年百货市场调查 (Competition Commission's Groceries Market Investigation 2007) 中的主要分析方法；而美国 1997 年 FTC v. Staples Depot 案，2012 年 Northshore University Health System 北岸大学健康系统案，以及 2013 年 In re Chocolate Confectionary Antitrust Litig. 巧克力糖果案中均使用了双重差分法。

间的长短不能由孩子自己决定，也不能由家长决定，否则那些天分高、自律程度高的孩子及家长可能会趋向于选择更长的受教育时间，而这在“随机实验”中需要加以避免。接下来，需要对这些被随机分配教育时长的孩子们进行观察，分析他们的收入水平会怎样受到教育时长的影响，这是最理想的实验状态，也就是所谓的“随机实验”设计。

但问题是，在经济学中并不能随便做这样的实验，在其他社会科学中也一样。首先存在伦理道德的限制。如果我们明明知道教育的诸多益处，凭什么随机选择缩短其中某些孩子的教育时长呢？为了解决这一问题，社会科学家们需要去寻找所谓的“自然实验”，也叫“准自然实验”。“自然实验”或“准自然实验”是指一个事件的发生或一个政策的执行导致了近似于真正的科学实验场景，而细心的研究人员能够发现这一场景，并借此来识别社会科学中的因果关系。在这个场景中，外生事件或政策（例子中的教育时长）是否作用于某个对象需要取决于某种随机成分，而体现这种随机成分的那个变量，通常就可以作为工具变量，可以理解为在实验中对政策等外生冲击进行随机分配这种作法的替代“工具”。

回到上面关于教育和收入关系的例子，获奖者的合作研究中使用了孩子的出生日期作为受教育时长的工具变量。为什么这是一个好的工具变量呢？这与上面讲到的政策实施的随机性有关。因为美国义务教育法规定，只要当年年满6岁的儿童都需要在当年9月份入学，而年满16岁生日之后才能合法地离开学校；结果造成同一年中出生日期较早的孩子会比日期较晚的孩子受教育时间更短。比如说有些州把每年12月31日作为入学年龄计算的截至日期，那么出生在第四季度的孩子不到6岁就可以上学了，而出生在第一季度的孩子大约要到6岁半才开始上学。所以，如果同样是过了16岁生日就离开学校，那么第四季度出生的孩子就会比第一季度出生的孩子教育年限更长！

在这个例子中，对学生入学年龄的要求和义务教育法的规定相结合，构成了一个“自然实验”，其中学生必须接受教育的年限跟他们的出生日期有关。这里，出生日期就可以作为教育年限的工具变量，因为孩子教育年限跟出生日期是密切

相关的，可以作为教育年限这一政策变量（模型中的原因变量）的分配工具；另一方面，因为教育年限是法定的，孩子的天分、自律性等因素对它的决定不起作用，所以这个工具变量就具有了随机的特征，就像是在实验室中随机地针对孩子们分配了教育年限，这就帮助我们实现了与“随机实验”设计相同的目的。

接下来可以把工具变量这个概念讲得更深入一些。什么样的变量是一个好的工具变量呢？这个变量需要符合两个条件：第一个条件是，这个变量要跟我们担心存在内生性也即可能受到其他混杂因素影响（或存在“共时性”问题）的那个变量有相关性。比如，孩子的出生日期是跟他受教育的时间长短直接相关的，因为根据美国的政策，必须要年满 16 周岁才可以选择辍学；第二个条件是，这个变量还应该跟左侧的被解释变量没有直接的关系。比如，孩子的出生日期应该跟他的收入水平没有直接的关系，也就是说孩子在哪个月份出生并不会直接影响他成年后的收入水平。前一个条件称为“相关性”，后一个条件称为“外生性”，满足以上两个条件的变量，就是一个好的工具变量。

讲到这儿，大家可能会松了一口气。工具变量法这种研究方法已经开发出来了，接下来的应用应该会比较容易了！但这种判断又为时过早了，因为在实践中应用这些方法仍然是非常具有挑战性的工作，过程中充满了艰辛！实际上，如何在实践中找到好的工具变量，被公认为是经济学研究中最困难的挑战之一。那么困难在哪里呢？且听下回分解！

**宋 健：**龙老师，您就经济分析中的因果识别以及工具变量法的介绍，听上去有些复杂和烧脑，但又很有趣，真的需要细细体会，我已经完全被吸引住了。非常期待您的下一期对谈！（待续）