

## 使用观察数据检验因果问题

### ——2021年诺贝尔经济学奖获得者的研究工作\*

海尔曼·佐尔坦，霍尔瓦特·海德维格，林德奈尔·奥蒂拉

上世纪80年代和90年代，劳动经济学处于结合经济理论、先进的计量经济学方法和新数据源的前沿。2021年诺贝尔经济学奖授予了在该研究计划中发挥关键作用的大卫·卡德（David Card）、约书亚·安格里斯特（Joshua Angrist）和吉多·伊本斯（Guido Imbens），以表彰他们“对劳动经济学的实证贡献”（Card）和“对因果关系分析的方法论贡献”（Angrist和Imbens）。这些方法论创新现在被用于经济学的所有领域和许多其他社会科学。由于真实的实证分析，经济学已经从一门理论性更强的学科转变为一门以实证结果为主导的科学，即使是最基本的经济理论也可以因为实证结果而被否定。在本文中，我们回顾了这一也被称为信誉革命时期的主要方法论成就，并通过一个个经济应用进行说明。

经济文献杂志（JEL）代码：J2，J6，I21，I26，C2，C26

关键词：自然实验；最低工资；教育回报；工具变量；差异中的差异；不连续回归

#### 一，前言

奥利·阿申菲尔特（Orley Ashenfelter）在1985年至2001年间担任最负盛名的经济学期刊之一《美国经济评论》（American Economic Review, AER）的主编，他是普林斯顿大学教授，也是三位获奖者中两位获奖者的博士生导师和实际上的共同作者，他在任期开始时惊奇地发现，发表在《新英格兰医学杂志》（New England Journal of Medicine, NEJM）这一医学期刊上的一篇论文（Hearst等1986）比《美国经济评论》的文章更能引起经济学家的共鸣。这篇论文研究了兵役对随后死亡率的影响，比较了越南战争期间入伍和未入伍的年龄组。由于对此感到惊讶，约书亚·安格里斯特开

---

\*所发表文章只代表作者本人的观点，不代表匈牙利国家银行的官方主张。

海尔曼·佐尔坦（Hermann Zoltán），匈牙利科学院经济与地区科学研究中心（KRTK）经济学研究所专家、布达佩斯考文纽斯大学副教授。电子邮箱：hermann.zoltan@krtk.hu

霍尔瓦特·海德维格（Horváth Hedvig），伦敦大学学院（University College London）助理教授。电子邮箱：h.horvath@ucl.ac.uk

林德奈尔·奥蒂拉（Lindner Attila），伦敦大学学院副教授。电子邮箱：a.lindner@ucl.ac.uk

本文大量借鉴了有关2021年诺贝尔经济学奖的公共和科学背景材料（瑞典皇家科学院2021 a,b）。

本文原文发表在《金融与经济评论》匈牙利语版2022年3月号。<https://doi.org/10.33893/FER.21.1.141>

始研究越南退伍军人工作市场前景，而艾伦·克鲁格（Alan Krueger）订阅了《新英格兰医学杂志》以找出其中发表的文章之所以成功的原因。他们与大卫·卡德发现，《新英格兰医学杂志》的文章清晰的方法论框架使分析变得透明，因而结果是可信的（Card 2021）<sup>1</sup>。由此，他们想到，有了透明的统计方法，经济研究也可以对公共政策产生更大的影响，同时学科的学术水平也会提高。这就是经济学中“分析框架”（research design）的建立方式，这一框架追随获奖者及其学界精英和其合著者的脚步，成为专业成熟度的新标准。

但是这个经济学中以前不为人知的概念，分析框架是什么意思？我们在下文中借助从获奖者作品中选出的例子来介绍这一点，这些例子清楚地说明了该学科的方法论发展，以及作为这场方法论革命的结果，我们对长期争论不休的经济问题的知识是如何发展的。

## 二，调查观测数据的因果问题：自然实验和分析框架

大多数应用研究的基本目标是揭示某种因果关系。在生命科学领域，最常用的方法是随机对照试验（RCT）。例如，在测试新药的有效成分时，将患者随机分配到处理组和对照组，通过比较两组的情况来衡量药物的因果效应。随机组成的两组成员的各种特征平均而言基本相同——因此处理后两组之间的差异只能用药物的作用来解释。在社会科学中，使用此类实验的可能性非常有限，但同时，在许多情况下我们可以找到可比较的处理组和对照组，我们并没有随机分配观察到的个体，而是最终结果还是类似的。越南战争期间，征兵是通过抽签进行的，所以谁当兵谁留在家里是随机的。其他时候，一个个公共政策改革也会导致类似的情况。当月出生的人与次月出生的人的义务教育或退休年龄不同。就居住在美国某个定居点的人而言，最低工资与几公里外的类似定居点不同，只是因为它属于另一个成员国。这些情况称为自然或准实验。如果我们在分析过程中识别并利用它们，即使没有随机对照实验，我们也可以令人信服地确定因果关系。大卫·卡德、约书亚·安格里斯特和圭多·伊本斯在这种方法的发展和传播中发挥了决定性作用，具有许多应用和分析方法的理论发展<sup>2</sup>。与以前的社会科学实践相比，这代表了方法的根本变化，后者试图根据两种现象的统计相关性来证明因果关系。以下是利用不同自然实验情况并使用不同分析方法的三个示例。从理论的角度来看，这三个案例都产生了重要的结果。

---

<sup>1</sup> Zipperer, B.: *Equitable Growth in Conversation: An interview with David Card and Alan Krueger*. (《对话中的公平增长：采访大卫·卡德和艾伦·克鲁格》) April, 2016. <https://davidcard.berkeley.edu/interviews/interview%20with%20Card%20and%20Krueger.pdf>. (下载日期：2022年2月22日)

<sup>2</sup> 应用的方法都是通过例子来展示的，所以下面我们依赖卡德和安格里斯特的文章。伊本斯在该方法论的发展过程中起到了决定性的作用，此处不再详述。

## 2.1. 我们对劳动力市场如何运作的了解

上世纪 80 年代和 90 年代，劳动经济学处于结合经济理论、先进的计量经济学方法和新数据源的前沿<sup>3</sup>。一个很好的例子就是关于最低工资的经济辩论，一些著名的经济学家[例如后来获得诺贝尔奖的乔治·斯蒂格勒（George Stigler）和米尔顿·弗里德曼（Milton Friedmann）]也对此发表了评论。

最低工资辩论基本上是关于什么是描述劳动力市场的正确理论。根据标准理论，低收入者的市场是竞争性的，因此工资上涨会导致失业（Stigler 1946）。在（完全）竞争的市场中，供需处于平衡状态。由于最低工资的引入，这种平衡被打破了，随着工资的提高，对劳动力的需求减少，因为它变得更加昂贵，而劳动力的供应增加，因为更多的人希望以更高的工资工作。这导致就业减少和失业。

与标准理论相反，许多人认为最低工资不会导致严重的失业。例如，根据理查德·莱斯特（Richard Lester 1947）的说法，标准理论存在缺陷，因为公司领导并不按照利润最大化公司的数学模型所描述的方式思考。他在问卷调查的帮助下支持了这一点，在问卷调查中，公司领导被问及是什么决定了他们雇用多少人。有趣的是，对于大多数公司管理者来说，最决定性的因素不是劳动力成本，而是对产品的需求。

然而，理查德莱斯特的问卷调查方法遭到了严厉的批评。诺贝尔奖获得者米尔顿·弗里德曼（Milton Friedman 1953）认为，竞争法则迫使利润最大化，即使企业领导没有通过求解数学方程式来实现利润最大化。根据该论点，作为竞争的结果，只有那些“感觉”到如何有效地经营公司和实现利润最大化的公司领导才会成功，即使他们不能像数学方程式那样清晰地表达这一点，也是如此。米尔顿弗里德曼用台球比赛的例子说明了这种方法：一个好的台球运动员不能用数学方程精确地描述在哪里击打台球，但是如果我们想了解台球的轨迹，我们可以这样做只有借助数学方程式。根据该论点，经济模型的检验应基于其预测（例如最低工资是否会减少就业），而不是基于假设（公司的劳动力需求是否由利润最大化决定）。

因此，关于最低工资的争论集中在检验标准理论的预测上。早期的实证结果表明，最低工资显著减少了就业（Brown 等 1982）。出于这个原因，到了上世纪 80 年代和 90 年代，绝大多数经济学家认为最低工资弊大于利。大卫·卡德和克鲁格（1994）的一项革命性研究打破了这一共识。

---

<sup>3</sup> *Interview with David Card.* (《采访大卫·卡德》) Federal Reserve Bank of Minneapolis, December 1, 2022. <https://www.minneapolisfed.org/article/2006/interview-with-david-card>. (下载日期：2022年2月22日)

早期的实证结果主要基于时间序列分析：人们检验了最低工资水平与失业或就业之间的时间联动性。这种方法的主要问题是最低工资水平不仅仅是偶然的结果，而是一个复杂的政治决策过程，甚至可能取决于经济周期。例如，如果在左翼政治领导下或在经济衰退期间更频繁地提高最低工资，则最低工资和就业的联动可能表明其他因素的影响，而不仅仅是最低工资。

为解决此类问题，卡德和克鲁格（1994）将所谓的“差异中的差异”估计方法应用于最低工资问题的实证分析。作者利用了美国的最低工资（部分）由成员国确定的事实。新泽西州于1992年4月提高了最低工资，而邻近的宾夕法尼亚州则没有。由于新泽西州和与新泽西州接壤的宾夕法尼亚州东部在许多方面非常相似，改革导致了一个准实验性的情况，即在特定地区（处理组）提高最低工资，同时越过州界，在另一个非常相似的区域没有增加（对照组）。

差分估计法将处理组（新泽西州）的就业变化与对照组（宾夕法尼亚州东部）的就业变化进行比较。由于宾夕法尼亚州东部的经济非常相似，在对照组的帮助下，我们可以过滤掉那些在没有最低工资增长的情况下可能发生在新泽西州的经济因素的影响，从而我们可以得到更准确的估计最低工资的因果效应。

卡德和克鲁格收集了新泽西州最低工资上涨前后新泽西州和宾夕法尼亚州东部快餐店就业情况的数据。使用这些结果表明，与对照组相比，处理组的就业增加了，这意味着最低工资没有减少，但增加了就业。因此，基于一种新的、统计上更可靠的分析框架的分析导致了与以前的主要是时间序列分析完全不同的结果，这些分析很难借助流行的经济理论来解释。

差分法已成为经济学方法论中使用最广泛的分析框架之一。该方法的优点是，由于个别公共政策对每个人的影响并不平等，因此经常会出现处理组和对照组的情况。实证方法基于以下假设：如果处理组未接受处理，则对照组和处理组的变化将相同。这就是所谓的“平行趋势”（parallel trends）假设，通常通过比较改革前的趋势来检验。卡德和克鲁格（2000）的文章就是一个很好的例子，其使用行政数据分析了1991年10月（最低工资上涨前6个月）至1995年10月（最低工资上涨后42个月）期间新泽西州和宾夕法尼亚州东部的就业趋势。结果如图1所示。

图1 新泽西州最低工资上涨的影响



可以清楚地看到，在最低工资上涨之前的六个月里，新泽西州和东宾夕法尼亚州的就业情况非常相似，而在1992年4月1日之后，新泽西州的就业人数比东宾夕法尼亚州略有增加。该图还表明，从长远来看，最初的积极影响不会变成消极影响：即使在最低工资上涨三年后，新泽西州的就业率仍高于宾夕法尼亚州东部。

差分法现在经常以改进形式使用，当没有与改革前的处理组相似的单一控制组时，但有多个控制观察（国家、成员国、公司、学校等）其中一个适当加权的组合很好地（“平行地”）符合被处理组的改革前趋势。圭多·伊本斯在这种“差异中的综合控制差异”方法的开发和统计改进中发挥了先驱作用（Arkhangelsky等2021；Athey–Imbens 2006，2022）。

尽管卡德和克鲁格1994年的分析受到严厉批评，但随后的研究结果支持了最初研究的主要发现<sup>4</sup>。多鲁克·曾吉兹（Doruk Cengiz）和他的合著者（2019）研究了138次大幅提高最低工资的影响，并得出结论认为，最低工资的提高不会对低收入者的就业产生负面影响。在总结了37项最低工资研究的结果后，保罗·沃尔夫森（Paul Wolfson）和戴尔·贝尔曼（Dale Belman）（Wolfson和Belman 2019）得出结论，最低工资对就业的影响非常小，在统计上与零没有区别。然而，这些结果必须始终在给定的背景下，在

<sup>4</sup>许多著名经济学家对卡德和克鲁格（1994）的分析感到愤怒，他们认为这种分析与基本经济理论相矛盾。例如，1986年获得诺贝尔经济学奖的詹姆斯布坎南就认为卡德和克鲁格的分析不科学，这样的文章在美国经济评论等主要经济学期刊上没有立足之地。

給定的最低工資水平下進行解釋。卡德和克魯格（1994）和隨後的研​​究都沒有聲稱最低工資可以提高到任何程度以上而不會失業。

值得注意的是，由於採用了新的、真實的實證分析方法，能夠否定在其之前獲得廣泛共識的經濟相關性。結果本身開創了一個重要的新研究方向，導致對勞動力市場的更現實的描述。這些新模型考慮到大多數公司不僅被動地接受市場工資，而且公司經常主動塑造它們。這種行為會導致勞動力市場的主導地位，其後果之一可能是最低工資在一定程度上增加了就業（Burdett – Mortensen 1998；Manning 2003）。此外，卡德和克魯格對最低工資的研​​究為所謂的信譽革命鋪平了道路，由此可以在“真實的”實證分析的幫助下，而不是在理論辯論中決定基本的經濟和社會政治問題。由於這一點，經濟學學科越來越多地轉向實證方向，增加了其科學有效性及其對公共政策的影響。

## 2.2. 教育研究的可信度革命

在教育研究中，以諾貝爾獎獲得者及其合著者的名字為標志的經濟學方法論轉變尤其具有革命性。在上世紀80年代之前，大多數社會學家、心理學家和其他社會科學學科的從業者都在研究教育政策主題。但是基於自然實驗情況的分析框架和它們所需的嚴格而透明的統計方法已經吸引了許多經濟學家關注這個話題，其中一個重要且相關的公共政策決策問題，決定了個人的生產力勞動力市場，等待答案。由於這種方法論的更新，即“真實性革命”（Angrist – Pischke 2010），教育經濟學在該領域的從業者數量和研究方法方面都達到了一個轉折點。在本節中，我們回顧了諾貝爾獎獲得者在該領域的開創性工作，重點關注兩組重要問題：教育的勞動力市場回報，以及學校/精英學校對學生成績的影響。

### 2.2.1. 教育的回報<sup>5</sup>

在經濟模型中，工資通常與員工的生產率密切相關。除了傳統的經濟話題，當經濟學家意識到教育可以顯著提高工人的生產力時，他們將注意力轉向了教育。教育被定義為具有回報/回報的人力資本投資。如何以及在多大程度上衡量這種回報是經濟學中一個基本且被廣泛討論的問題。自從雅各布·明瑟（Jacob Mincer）（1958，1974）進行研究這方面工作以來，所謂的基於觀測數據的人力資本收益方程的估計，也被稱為勞動經濟學家先驅明瑟方程。該方程的典型形式為：

$$\log y = a + bS + cX + dX^2 + e,$$

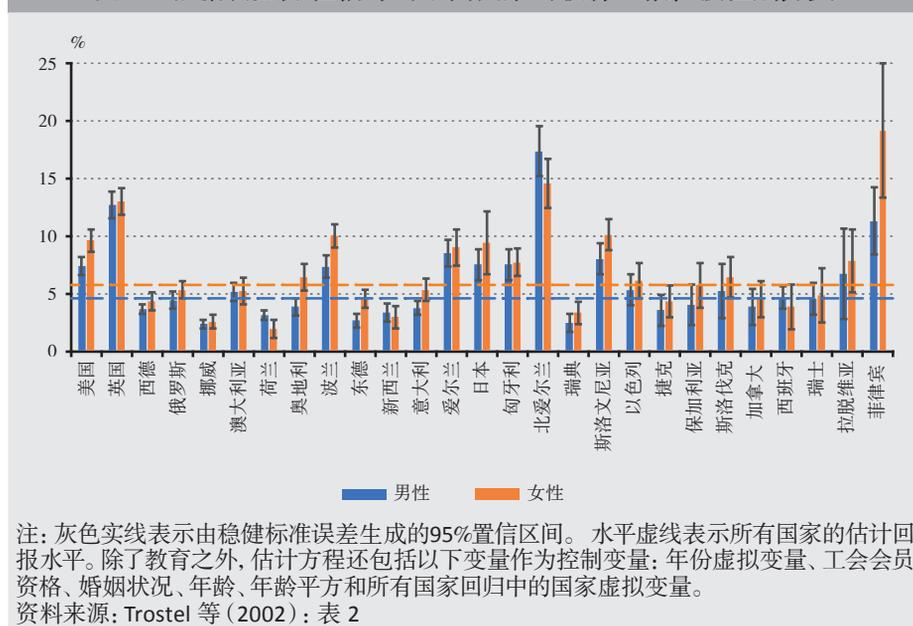
其中  $\log y$  是收入的对数， $S$  是教育程度（以毕业或在校学习的年数衡量）， $X$  是劳动力市场经验，即已工作年数<sup>6</sup>， $e$  是残差项，其中包含等式统计估计形式的附加控制变量（例如性别、婚姻状况、工会会员资格等）。

<sup>5</sup> 本节主要借鉴了卡德（1999）的观点。

<sup>6</sup>  $X$  实际上很难被观察到，所以通常用潜在的劳动力市场经验来近似： $A - c - S$ ，其中  $A$  是个体的年龄， $c$  是义务教育的最低年龄。

尽管明瑟从个人教育选择模型中得出这一点，并且适合数据的形式——例如，控制变量的范围或教育的函数形式，潜在的劳动力市场经验——可以改变，但经验关系的程度（ $b$ ）在许多情况下，使用相同的估计程序[使用普通最小二乘法（ordinary least squares, OLS）]具有相似的结果：加上在学校度过的一年大约。它与以后 4–10% 的高收入相关（图 2）。

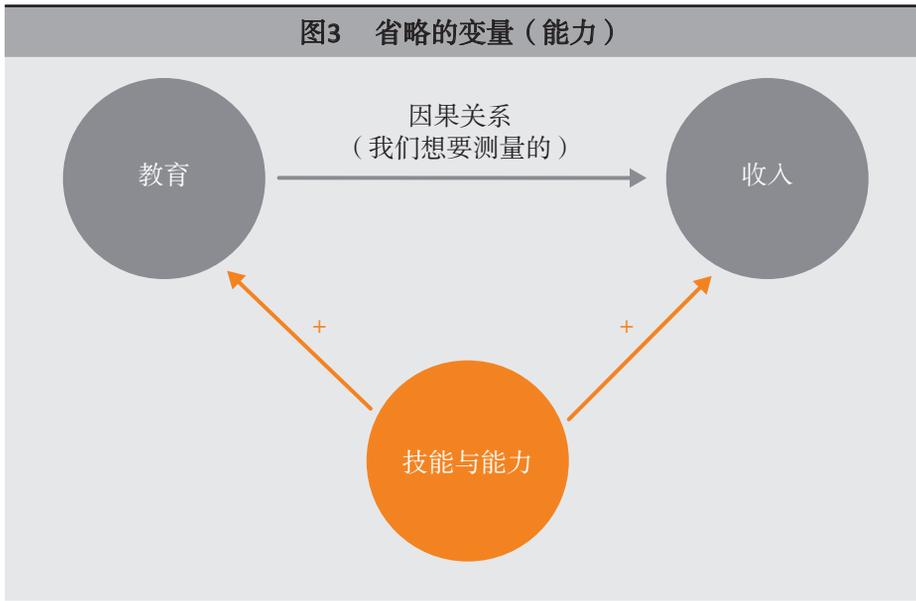
图2 根据明瑟方程估计的不同国家的教育回报，按性别分类



我们经常可以听到这种说法——误导性地——措辞为“在学校多度过一年，以后就会多增加 4–10% 的收入”。“增加”一词可能会给我们一种印象，即高等教育会带来更高的收入。部分原因可能是因为我们很有可能对这种因果关系感兴趣。例如，我们想知道我们的孩子是否应该上大学。或者，作为公共政策顾问，我们必须回答是否值得支持扩大中等/高等教育的问题。然而，由于多种原因，使用 OLS 估计的回归结果不适合得出这样的因果结论。

首先，由于很难观察到个人的技能和能力，所以我们通常没有关于这方面的数据。然而，能力较好的孩子可能会有更好的学习成绩，因此他们更有可能学习更长的时间（例如上大学）。另一方面，更好的技能/能力直接影响收入，即使没有受过高等教育也会增加收入<sup>7</sup>。因此，当我们在不观察能力的情况下发现教育与收入之间存在正相关的关系时，我们无法确定我们实际上看到的是教育的影响还是更好能力的影响。图 3 对此进行了说明。

<sup>7</sup> 这些变量，我们要么没有数据，要么我们“忘记”将它们包括在回归模型中，即使它们与我们的主要自变量（教育）和我们的输出变量（收入）相关，也被称为遗漏变量。



从明瑟方程估计的系数结合了两种影响，即教育的直接因果影响和更好能力的间接影响（能力偏差）。由于我们预计两者都是积极的，我们说可能高估了我们感兴趣的教育的纯因果效应。

将明瑟的教育系数解释为因果效应的第二个问题是，如果个人之间在多读一年的学校回报有多大方面存在很大差异，那么这些人可能会决定（“自我选择”）完成高水平的教育，教育为他们带来特别高的回报。换句话说，可以想象存在反向因果关系，这也加强了教育与收入之间的正向因果关系。因此，我们只能预期明瑟方程中的OLS估计是向上偏的。

2021年诺贝尔经济学奖得主们恰恰强调了相关性并不一定意味着因果关系，并开发了方法使其类似于随机实验，但在现实生活中自然发生，用获奖者及其合著者创造的表达方式来说——我们可以根据“自然实验”情况下产生的观察数据来衡量因果效应。在这些准实验分析框架中，除了前面提到的“差异中的差异”方法之外，另一个，也许是最广为人知的，就是所谓的工具变量（instrumental variables, IV）方法，该方法在1990年代经常用于衡量教育回报。

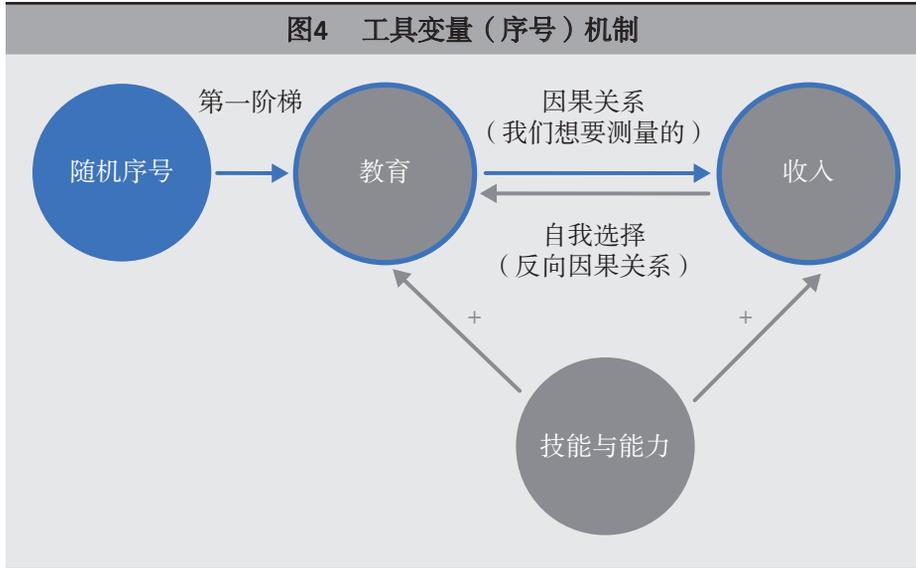
通过安格里斯特和克鲁格（1992）的例子，图4中的蓝色部分说明了IV框架的本质。该研究使用该想法作为分析框架，即在美国的越南战争期间，那些上大学的人可以避免征兵。同时，征兵的方式是，每名军龄男子根据出生年月随机获得一个编号，然后根据军队的人力需求，按编号从小到大的顺序征召<sup>8</sup>。在这种特定的监管环境中，用于测序的序列号可以被认为是一个工具变量，其

<sup>8</sup> 本文开始提到的诺曼·赫斯特（Norman Hearst）等（1986）以及由此受到启发的安格里斯特（1990）使用了与此相同的排序顺序。

1, 可以确定教育水平——因为可以通过入读大学来避免征兵, 纯属偶然, 序列号较低的人 (较早应征入伍的人) 更有可能上大学,

2, 在教育以外的渠道, 例如它不会影响通过技能获得的收入——同样, 因为序列号是随机分配的 (基于与技能无关的出生日期)。

也就是说, 工具的变化 (较低的等级) 在不改变能力的情况下增加了教育。因此, 个人之间的任何收入差异只能追溯到序列号的差异, 从而导致教育程度的差异, 而不是能力的差异。在这种情况下, 我们可以确定收入的差异是由教育的差异引起的, 因为能力不会因序列号的影响而有所差异。



一般而言, IV方法的巨大优势在于, “好的”工具能够从遗漏变量 (教育回报情况下的技能) 和反向因果关系 (自我选择) 中清除OLS估计值<sup>9</sup>。这种方法的缺点是很难找到“好的”工具, 因为它必须满足以下两个条件——说得更严格一点:

1. 相关性: 工具必须与我们感兴趣的内生变量 (在我们的例子中是教育) 密切相关, 但由于上述遗漏变量/反向因果关系问题, 我们用OLS只会得到扭曲的估计。

2. 排他性: 工具不得通过任何其他渠道与输出变量 (在我们的示例中为收益) 相关。

为了确保相关性, 我们使用所谓的第一阶段回归, 对工具和其他控制变量 (例如潜在的劳动力市场经验) 的内生变量 (教育) 进行回归。在这个回归中, 我们可以根据 F 统计量测试工具系数的联合显著性来确定工具的相关

<sup>9</sup> 甚至, 它还可能解决本文未单独讨论的教育回报测度中的第三个问题: 教育测度误差导致的偏差。有关详细信息, 请参见卡德 (1999)。

性，即工具是否強大<sup>10</sup>。儘管有關於我們的工具在 F 值之上多少足夠強大/相關的經驗法則（Staiger – Stock 1997；Stock – Yogo 2005），事實上，如果有人不完全相信工具的排他性（有效性），就不可能毫無疑問地確定這一點。即使只是一點點損害排他性的弱工具（也通過其他渠道影響收入，而不僅僅是教育），都甚至可能增加 OLS 估計的偏差（稍後詳見）。然而，特別難以確信排他性的實現，因為沒有正式的統計測試。因此，對於研究人員來說，唯一的可能性就是部分地、準確地支持它，使工具的假定操作透明化，即明確“識別從何而來”。

卡德（1995）的文章<sup>11</sup>可能是受自傳启发，將個人出生地與最近大學的距離用作教育工具（大學學歷）。這是一個自然實驗，因為個人的出生地可以被認為是隨機的，但如果離大學很近，個人上大學的成本就會比離得遠的地方低。出於這個原因，那些出生在附近的人比那些出生在較遠的地方的人更有可能從大學入學和畢業——該論點支持該工具的相关性。然而，誰出生在大學附近與個人的收入沒有任何其他關係 - 支持排他性的論點說。然而，事實證明，沒有一個條件是完全有效的：工具只在教育背景較低的人中很強，而那些靠近大學的地方通常會提供更好的工作機會，即在另一個地方靠近大學方式——通過更有利的當地勞動力市場機會——也會影響收入。與本文中衡量教育回報的許多其他 IV 研究一樣，數值結果表明 IV 估計值遠高於 OLS 估計值。從那以後，這一直是一個未解之謎，因為能力引起的遺漏變量問題和自我選擇導致的反向因果關係都會預設 OLS 向上偏，因此無偏 IV 估計會導致較小的回報。對這種明顯的矛盾有什麼可能的解釋？

一方面，工具可能無效（即不滿足排他性）和弱，這種組合可能會放大 OLS 估計中的偏差。研究人員認為這不可能/令人信服地充分解釋了遠高於 OLS 的 IV（Card 1999）。例如，在卡德（1995）的情況下，即使我們控制個人的家庭背景，IV 也比 OLS 大 30%，此時工具的排他性和相关性更有說服力。

另一方面，可以想象 IV 更大，因為工具還消除了由測量誤差引起的 OLS 的向下偏差（見腳注 9）。

第三，根據包括卡德（1999）在內的許多人的說法，IV 研究很可能估計了處理對特殊群體的影響（在我們的例子中為教育回報），對他們來說，這大於估計的平均效應給大家。為了更好地闡明我們的意思，請考慮安格里斯特和克魯格（1991）對教育回報的估計。在此，作者利用了美國公共教育系統的特殊性，即在大多數州，在給定年份的 9 月 1 日，只有那些在該學年的 1 月 1 日之前年滿 6 歲的人開始接受小學教育，而其他人則只一年後開始。出於這個原因，1 月 1 日（第一季度）出生的人通常在 6 又 3/4 歲開始上學，而 12 月 31 日（第四季度）出生的人則年輕一歲，在 5 又 1/2 歲開始上學。因此，到 16 歲生日，即達到義務教育年齡上限時，一季度出生的人比四季度出生的人少

<sup>10</sup> 在前面提到的安格里斯特和克魯格（1992）的文章中，發現序列號作為一種工具很弱——序列號低與一個人是否上過大學沒有很强的相关性（Card 1999）。

<sup>11</sup> 卡德在加拿大的一个农场长大，但附近有一所大学。他也将自己的教育归因于此。（《与大卫·卡德的访谈》，明尼阿波利斯联邦储备银行，2006年12月1日。<https://www.minneapolisfed.org/article/2006/interview-with-david-card>）

上一年学（10年多一点）。如果出生季度独立于决定收入的其他因素，则可以将其用作教育工具，第一季度和第四季度出生的人的收入差异表明了教育差异的产出。然而，正如安格里斯特和克鲁格所指出的，我们只看到早期离校生的教育存在差异，而在继续接受大学或研究生培训的人中则没有。收入的差异也集中在这个群体。然而，这是一个特殊的群体：他们是仅靠义务教育而在校时间较长的。这项研究衡量的是这一特殊群体的教育回报，而不是普通人群。然而，这一群体的教育回报可能与普通人群大不相同。

这个想法被伊本斯和安格里斯特（1994）以及安格里斯特等人正式化（1996）。所谓在局部平均处理效应（local average treatment effect）的概念中。仍然使用安格里斯特和克鲁格（1991）的例子，如果出生在第一季度和第四季度，人口可以根据个人上学/将上学的多少分为4个部分<sup>12</sup>：

- 1，始终参与处理者（always takers）：无论第一季度还是第四季度出生，都始终上学；
- 2，从不参与处理者（never takers）：无论第一季度还是第四季度出生，上学时间很短；
- 3，遵循规定者（compliers）：如果是第四季度出生，与第一季度出生的一样继续上学；
- 4，违反规定者（defiers）：如果是第一季度出生，与第四季度出生的一样继续上学。

伊本斯、安格里斯特和他们的合著者表明，在一定的条件下——例如，如果没有违反规定者——IV预估会预估处理（在我们的例子中是上学接受教育）对遵循规定者群体的影响<sup>13</sup>。然而，在不同工具的情况下，遵循规则者的群体可能会不同，对他们来说，教育的回报可能不同。例如，安格里斯特和克鲁格（1991）发现，在接受义务教育而留在高中的组中，额外一年教育的回报约为7.5%，这与OLS估计值相差无几。相比之下，卡德（1995）发现，在遵循规则者群体中（那些只在离他们很近的情况下才上大学的人），教育的回报——在他们的例子中是大学——远高于OLS估计的回报。

安格里斯特和克鲁格（1991）的文章被广泛重新审视。约翰·邦德（John Bound）等（1995）表明，四分之一出生作为一种工具的相关性很弱，因此即使样本很大，IV估计值也可能不一致。这篇文章开启了整个方法论研究的浪潮，在研究人员有许多弱工具可供使用的情况下提出实用建议（例如Staiger – Stock 1997）。今天，随着机器学习方法的传播，正在为这个问题创造新的解决方案<sup>14</sup>。

<sup>12</sup> 请注意，从数据中看不到谁属于哪个组，因为划分不仅基于在实际生日的情况下选择的教育，而且还基于所谓的也用于反事实选择，即给定的个人如果出生在不同的时间会选择什么。后者是不可观察的。然而，借助伊本斯和安格里斯特开发的方法，我们可以用数据中可以观察到的特征来描述这些群体。

<sup>13</sup> 此外，作者还详细阐述了如何根据数据中观察到的特征来描述规则遵循者群体的特征。

<sup>14</sup> 例如，埃洛拉·德伦库特（Ellora Derenoncourt）（2022）的文章中对亚历山大·贝洛尼（Alexandre Belloni）等（2011）所写文章的采用。

### 2.2.2. 名校效應

對於教育研究人員、教育政治家和家長來說，這是一個基本問題，哪些學校是好學校，是什麼讓他們比其他學校做得更好。乍一看，這似乎是一個微不足道的問題：只需看看哪些學校在所有學校排名中名列前茅即可。這些排名基於非常清楚地顯示學校學生表現的數據：高中畢業率、高等教育率、標準化考試成績。在大多數國家，學校之間在這些指標方面存在顯著差異，通常會有少數精英學校脫穎而出。

同時，也很明顯的是，這些學校的學生構成也大不相同。因為它們很受歡迎，所以它們可以從眾多申請者中選出表現最好的學生，然後真正表現出色的學生。但選拔和學校的作用有多大，教育質量更高？如果政府要鼓勵學校提高教育質量，這個問題就顯得尤為重要，因為這需要衡量質量。在美國，1990年代有許多成員國和大都市地區都推出了這種教育改革，並且在2000年代初期，這也以聯邦計劃的形式發表（No Child Left Behind program）。

長期以來，研究人員一直在嘗試使用各種統計方法來分離學生特徵和學校質量的影響。一方面，我們可以嘗試直接過濾掉學生特徵的影響，將他們作為控制變量納入分析。另一方面，我們可以檢查個別學生的表現隨時間的變化（附加值）。這種方法是基於這樣的假設，即學生個人特徵的影響被總結在先前取得的考試成績中，因此他們考試成績的變化或多或少可以歸因於學校。這兩種方法，或它們的各種組合，經常被用來衡量個別學校或個別教師的有效性。大家都一致認為，與輸出指標的簡單平均數相比，這些估計通常更接近學校的實際質量，但它們的準確性如何是一個懸而未決的問題。有充分的理由認為，在某些情況下，例如優秀的精英學校，學生的不可觀察特徵的作用可能非常大，這是上述傳統方法無法考慮的。申請這些學校的學生想必更上進、更勤奮，他們的学习成績對他們的父母來說極其重要，他們在廣義上對他們的发展投入更多。想必這些學生發展得更快，所以相比於他們之前的成績提升幅度更大，所以即便是計算附加值也不能完全排除選拔的影響。

如何更準確地衡量學校的質量？如果把學校質量看成是學校對學生知識和技能的影響，我們可以運用因果分析的方法和自然實驗的邏輯。安格里斯特的研究在這一領域起到了決定性的作用。

他及其合著者在幾項研究中調查的問題之一——正如他在諾貝爾獎頒獎典禮上的演講中強調的那樣——長期以來一直困擾着他的，就是所謂的特許學校的有效性。這些是非公立學校，作為與政府簽訂的合同（章程）的一部分獲得國家資助（在美國，與匈牙利的做法相反，這通常不適用於私立和教會學校），他們可以執行不同於傳統公立學校的課程和教學實踐。它們經常在隔離的大都市地區開展業務，並成功地為居住在那裡的貧困學生提供教育。許多人在這些學校中看到了更新美國教育體系的机会，並期望它們能夠縮小黑人和白人學生之間巨大的成績差距，而另一些人則表示，選拔也是特許學校成功的背後原因；申請這裡的學生和他們的父母比來自類似社會情況的學生的平均水平更有動力和投入。

安格里斯特和他的合著者研究了马萨诸塞州的特许学校，在超额报名的情况下，会使用抽签来决定谁能进入该学校（Angrist等 2010；Angrist等 2012）。他们将抽签作为一项自然实验，将幸运者和被排除在外的人后来的结果进行比较。使用IV方法，将抽签的结果作为工具。就被调查学校而言，显示出显著的积极影响，录取学生的考试成绩提高了。例如，一年之后，与那些因同龄人抽签而未被录取的人相比，数学提高了标准差单位的三分之一以上，阅读理解提高了十分之一标准差单位（Angrist等 2012）<sup>15</sup>。

安格里斯特和他的同事取得重大成果的另一个问题是对精英学校影响的调查（Abdulkadiroğlu等 2014）。他们在开创性的工作中研究了波士顿和纽约各前三名精英高中的影响。学生根据一次入学考试的结果被录取到这些学校。直接比较被录取和被拒的申请者并不能看出学校的效应，因为这两个群体有很大的不同：被录取的学生不是随机录取的，他们过去的学习成绩要好得多。同时，录取门槛两边的刚刚能被录取的与落后学生的情况则很相似；我们可以假设他们在未观察到的特征上完全没有区别。当然，录取门槛以下的学生之前的表现有些弱，我们在分析时必须考虑到这一点。在这种情况下，当被处理组的概率突然显著增加到超过给定变量的某个值时，我们可以应用回归断点设计方法（regression discontinuity design），就像安格里斯特和他的同事们在分析精英学校时所做的那样。他们的研究不仅对学校的影响分析做出了贡献，也成为了应用间断回归法作为分析框架的范例。

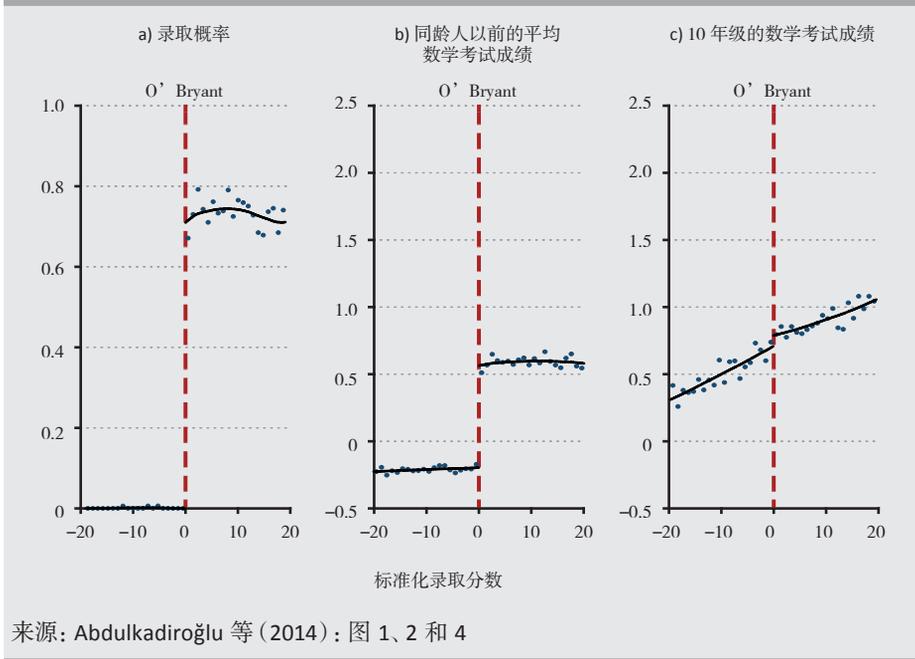
该方法的出发点是，在处理组中的位置取决于特定变量（*running*或*forcing variable*）的值，例如，如果学生的分数超过某个阈值，他们就可以进入精英学校。录取期间提供的表现也与后来的结果有关，用拟合两个变量的回归曲线表示（可以使用各种参数或非参数方法进行估计）。被置于处理组中的效应通过在输出变量的阈值处观察到的“跳跃”的幅度来显示。

图 5 说明了波士顿一所精英学校的结果。可以清楚地看到，在入学门槛处，入学概率急剧增加（图5a），同龄人的构成也发生了显著变化（图5b）。同时，出乎我们意料的是，学生后期的数学考试成绩并没有大幅提升（图5c）：进入名校的学生和刚错过名校的学生后期的学习成绩并没有彼此不同。这一令人惊讶的结果也体现在研究的标题中：精英错觉。

---

<sup>15</sup> 相比之下，这些都是很大的影响；三年级黑人和白人学生的原始考试成绩差异几乎是 1 个标准差（Fryer – Levitt 2006）。

图5 精英错觉



从几个角度来看, 精英学校对考试成绩的零影响很有趣。首先, 它有助于教育中研究同学影响的文献。班级的构成如何影响一个特定的学生, 他的同学是好学生还是无纪律的学生? 关于这个问题的文献提供了许多直接不同的结果, 但这种影响在实证分析中很难识别 (Angrist 2014)。安格里斯特和他的同事们的研究间接回答了这个问题。说明进入名校学生的同学以往平均考试成绩远高于退学学生, 但这对给定学生后来的成绩没有影响, 即对本例中的成绩同学是无所谓的<sup>16</sup>。

同时, 问题来了, 如果没有实际效应, 为什么这些学校会如此受到家长和学生的欢迎呢? 申请者追逐精英教育的幻想会错吗? 还是以考试成绩衡量的学生成绩并不能体现学生在这些学校上学后的收获? 这些暂时是悬而未决的问题, 但最近一项分析芝加哥精英学校的研究表明这些偏好有多强烈 (Angrist等 2019)。在这里, 作者发现, 被精英学校录取会对学生的成绩产生负面影响, 因为许多未被录取的学生所上的特许学校可以提高他们的成绩。然而, 申请人更喜欢精英学校而不是特许学校。

<sup>16</sup> 安格里斯特和凯文·朗 (Kevin Lang) (2004) 在之前的一项研究中得出了类似的结果: 在波士顿的一个融合计划框架内, 花园城市学校中产阶级学生的表现不受一些市中心、贫困、班上的黑人学生。安格里斯特 (2014) 认为, 许多分析高估了同伴的影响力, 而且它实际上比我们想象的要弱得多。

只有在特殊情况下才能对学校的影响进行因果估计，并且结果涉及特定的，通常是特定的学生群体和所涉及的学校。我们通常可以从此类分析中学到什么？在最近的一项研究中（Angrist等 2017），安格里斯特和他的合著者利用了这样一个事实，即在波士顿学校系统中，学生通过中央招生匹配算法分配到六年级的高中，并随机抽取算法中起着重要作用<sup>17</sup>。通过这种方式，他们能够估计学校对许多学校的因果效应，并将其与传统的学校效能增值指标进行比较<sup>18</sup>。人们发现，使用增值法估算的学校质量确实给出了有偏差的估计，但这种偏差并没有大到教育政策制定者不值得依赖这些数据来做决定。

### 三、结语

基于自然实验的新方法的出现从根本上改变了经济研究，约书亚·安格里斯特和约恩-斯特芬·皮什克（Jörn-Steffen Pischke）（2010）确切地将这种变化称为“可信度革命”。这需要开发新的分析方法，其中大卫·卡德、约书亚·安格里斯特和圭多·伊本斯发挥了决定性作用。这使得经济学更接近理解和实证验证因果关系的自然科学理想。但也许更重要的是，这个新工具箱有助于更好地理解社会和经济背景。

### 参考文献

- Abdulkadiroğlu, A. – Angrist, J. – Pathak, P. (2014): *The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.* (《精英幻觉：波士顿和纽约考试学校的成就效应》) *Econometrica*, 82(1): 137–196. <https://doi.org/10.3982/ECTA10266>
- Angrist, J.D. (1990): *Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.* (《终生收入和越南时代的彩票草案：来自社会保障行政记录的证据》) *American Economic Review*, 80(3): 313–336.
- Angrist, J.D. (2014): *The perils of peer effects.* (《同伴效应的危险》) *Labour Economics*, 30(October): 98–108. <https://doi.org/10.1016/j.labeco.2014.05.008>
- Angrist, J.D. – Dynarski, S.M. – Kane, T.J. – Pathak, P.A. – Walters, C.R. (2010): *Inputs and Impacts in Charter Schools: KIPP Lynn.* (《特许学校的投入和影响：KIPP Lynn》) *American Economic Review*, 100(2): 239–243. <https://doi.org/10.1257/aer.100.2.239>
- Angrist, J.D. – Dynarski, S.M. – Kane, T.J. – Pathak, P.A. – Walters, C.R. (2012): *Who Benefits from KIPP?* (《谁从 KIPP 中受益》) *Journal of Policy Analysis and Management*, 31(4): 837–860. <https://doi.org/10.1002/pam.21647>

<sup>17</sup> 在匈牙利，类似的中央录取算法将申请人与中学和大学的学校相匹配，但随机因素在这方面的作用可以忽略不计。

<sup>18</sup> 在这样做时，他们效仿了罗伯特·拉隆德（Robert LaLonde）（1986）关于劳动力市场培训计划影响的开创性研究的例子。

- Angrist, J.D. – Hull, P.D. – Pathak, P.A. – Walters, C.R. (2017): *Leveraging Lotteries for School Value-Added: Testing and Estimation*. (《利用彩票促进学校增值: 测试和估算》) *Quarterly Journal of Economics*, 132(2): 871–919. <https://doi.org/10.1093/qje/qjx001>
- Angrist, J.D. – Imbens, G.W. – Rubin, D.B. (1996): *Identification of Causal Effects Using Instrumental Variables*. (《使用工具变量识别因果效应》) *Journal of the American Statistical Association*, 91(434): 444–455. <https://doi.org/10.2307/2291629>
- Angrist, J.D. – Krueger, A.B. (1991): *Does compulsory school attendance affect schooling and earnings*. (《义务教育是否影响上学和收入》) *Quarterly Journal of Economics*, 106(4): 979–1014. <https://doi.org/10.2307/2937954>
- Angrist, J.D. – Krueger, A.B. (1992): *Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery*. (《使用越南时代的抽签来估计学校教育的收益》) NBER Working Paper No. 4067. <https://doi.org/10.3386/w4067>
- Angrist, J.D. – Lang, K. (2004): *Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program*. (《学校融合会产生同伴效应吗? 来自波士顿 Metco 计划的证据》) *American Economic Review*, 94(5): 1613–1634. <https://doi.org/10.1257/0002828043052169>
- Angrist, J.D. – Pathak, P.A. – Zárate, R.A. (2019): *Choice and Consequence: Assessing Mismatch at Chicago Exam Schools*. (《选择与后果: 评估芝加哥考试学校的不匹配》) NBER Working Paper No. 26137. <https://doi.org/10.3386/w26137>
- Angrist, J.D. – Pischke, J.-S. (2010): *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics*. (《实证经济学中的可信度革命: 更好的研究设计如何消除计量经济学中的骗局》) *Journal of Economic Perspectives*, 24(2): 3–30. <https://doi.org/10.1257/jep.24.2.3>
- Arkhangelsky, D. – Athey, S. – Hirshberg, D.A. – Imbens, G.W. – Wager, S. (2021): *Synthetic Difference-in-Differences*. (《综合差异中的差异》) *American Economic Review*, 111(12): 4088–4118. <https://doi.org/10.1257/aer.20190159>
- Athey, S. – Imbens, G.W. (2006): *Identification and Inference in Nonlinear Difference-in-Differences Models*. (《非线性差分模型中的识别和推理》) *Econometrica*, 74(2): 431–497. <https://doi.org/10.1111/j.1468-0262.2006.00668.x>
- Athey, S. – Imbens, G.W. (2022): *Design-based analysis in Difference-In-Differences settings with staggered adoption*. (《交错采用的差异设置中基于设计的分析》) *Journal of Econometrics*, 226(1): 62–79. <https://doi.org/10.1016/j.jeconom.2020.10.012>
- Belloni, A. – Chernozhukov, V. – Hansen, C. (2011): *LASSO Methods for Gaussian Instrumental Variables Models*. (《高斯工具变量模型的 LASSO 方法》) arXiv:1012.1297v2. <https://arxiv.org/abs/1012.1297>

- Bound, J. – Jaeger, D.A. – Baker, R.M. (1995): *Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak*. (《当工具与内生解释变量之间的相关性较弱时, 工具变量估计的问题》) *Journal of the American Statistical Association*, 90(430): 443–450. <https://doi.org/10.1080/01621459.1995.10476536>
- Burdett, K. – Mortensen, D.T. (1998): *Wage differentials, employer size, and unemployment*. (《工资差异、雇主规模和失业率》) *International Economic Review*, 39(2): 257–273.
- Brown, C. – Gilroy, C. – Kohen, A. (1982): *The effect of the minimum wage on employment and unemployment*. (《最低工资对就业和失业的影响》) *Journal of Economic Literature*, 20(2): 487–528.
- Card, D. (1995): *Using geographic variation in college proximity to estimate the return to schooling*. (《使用大学邻近度的地理差异来估计返校率》) In: Christofides, L.N. – Grant, E.K. – Swidinsky, R. (eds.): *Aspects of labour market behaviour: Essays in honour of John Vanderkamp*. University of Toronto Press, Toronto, Canada, pp. 201–222.
- Card, D. (1999): *The Causal Effect of Education on Earnings*. (《教育对收入的因果影响》) In: Ashenfelter, O.C. – Card, D. (eds.): *Handbook of Labor Economics*. Elsevier, Volume 3, Part A, pp. 1801–1863. [https://doi.org/10.1016/S1573-4463\(99\)03011-4](https://doi.org/10.1016/S1573-4463(99)03011-4)
- Card, D. (2021): *Design-Based Research in Empirical Microeconomics*. (《实证微观经济学中基于设计的研究》) Nobel Memorial Lecture, November 2021. <https://davidcard.berkeley.edu/papers/CardNobelLecture.pdf>
- Card, D. – Krueger, A. (1993): *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*. (《最低工资和就业: 新泽西州和宾夕法尼亚州快餐业的案例研究》) *American Economic Review*, 84(4): 772–793.
- Card, D. – Krueger, A. (2000): *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply*. (《最低工资和就业: 新泽西州和宾夕法尼亚州快餐业的案例研究: 回应》) *American Economic Review*, 90(5): 1397–1420. <https://doi.org/10.1257/aer.90.5.1397>
- Cengiz, D. – Dube, A. – Lindner, A. – Zipperer, B. (2019): *The effect of minimum wages on low-wage jobs*. (《最低工资对低薪工作的影响》) *Quarterly Journal of Economics*, 134(3): 1405–1454. <https://doi.org/10.1093/qje/qjz014>
- Derenoncourt, E. (2022): *Can You Move to Opportunity? Evidence from the Great Migration*. (《你能抓住机会吗? 来自大迁徙的证据》) *American Economic Review*, 112(2): 369–408. <https://doi.org/10.1257/aer.20200002>
- Friedman, M. (1953): *The Methodology of Positive Economics*. (《实证经济学方法论》) In: *Essays in Positive Economics*, University of Chicago Press, pp. 3–43.

- Fryer, R.G. – Levitt, S.D. (2006): *The Black-White Test Score Gap Through Third Grade*. (《三年级的黑白测试分数差距》) *American Law and Economics Review*, 8(2): 249–281. <https://doi.org/10.1093/aler/ahl003>
- Hearst, N. – Newman, T.B. – Hulley, S.B. (1986): *Delayed effects of the military draft on mortality: A randomized natural experiment*. (《征兵对死亡率的延迟影响：一项随机自然实验》) *New England Journal of Medicine*, 314(10): 620–624. <https://doi.org/10.1056/NEJM198603063141005>
- Imbens, G.W. – Angrist, J.D. (1994): *Identification and Estimation of Local Average Treatment Effects*. (《局部平均处理效应的识别和估计》) *Econometrica*, 62(2): 467–475. <https://doi.org/10.2307/2951620>
- LaLonde, R.J. (1986): *Evaluating the Econometric Evaluations of Training Programs with Experimental Data*. (《用实验数据评估培训计划的计量经济学评价》) *American Economic Review*, 76(4): 604–620.
- Lester, R.A. (1947): *Marginalism, Minimum Wages, and Labor Markets*. (《边际主义、最低工资和劳动力市场》) *American Economic Review*, 37(1): 135–148.
- Manning, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*. (《运动中的垄断：劳动力市场中的不完全竞争》) (Course Book ed.). Princeton: Princeton University Press. <https://doi.org/10.1515/9781400850679>
- Mincer, J. (1958): *Investment in Human Capital and Personal Income Distribution*. (《人力资本投资与个人收入分配》) *Journal of Political Economy*, 66(4): 281–302. <https://doi.org/10.1086/258055>
- Mincer, J. (1974): *Schooling, experience and earnings*. (《学校教育、经验和收入》) Columbia University Press, New York.
- Royal Swedish Academy of Sciences (2021a): *Natural experiments help answer important questions*. (《自然实验有助于回答重要问题》) An easy-to-understand background to the 2021 Nobel Prize in Economics. <https://www.nobelprize.org/uploads/2021/10/popular-economicsciencesprize2021-3.pdf>. (下载日期：2022年1月10日)
- Royal Swedish Academy of Sciences (2021b): *Answering causal questions using observational data*. (《使用观察数据回答因果问题》) A scientific background to the 2021 Nobel Prize in Economics. <https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf>. (下载日期：2022年1月10日)
- Staiger, D. – Stock, J.H. (1997): *Instrumental Variables Regression with Weak Instruments*. (《使用弱工具的工具变量回归》) *Econometrica*, 65(3): 557–586. <https://doi.org/10.2307/2171753>
- Stigler, G.J. (1946): *The Economics of Minimum Wage Legislation*. (《最低工资立法的经济学》) *American Economic Review*, 36(3): 358–365.

- Stock, J.H. – Yogo, M. (2005): *Testing for weak instruments in Linear IV Regression*. (《测试线性IV回归中的弱工具》) In: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press, pp. 80–108. <https://doi.org/10.1017/CBO9780511614491.006>
- Trostel, P. – Walker, I. – Woolley, P. (2002): *Estimates of the economic return to schooling for 28 countries*. (《对 28 个国家的教育经济回报的估计》) *Labour Economics*, 9(1): 1–16. [https://doi.org/10.1016/S0927-5371\(01\)00052-5](https://doi.org/10.1016/S0927-5371(01)00052-5)
- Wolfson, P. – Belman, D. (2019): *15 years of research on U.S. employment and the minimum wage*. (《15年美国就业与最低工资研究》) *Labour*, 33(4): 488–506. <https://doi.org/10.1111/labr.12162>