

实证研究中的语言逻辑与数理逻辑

姓 名：陈硕

单 位：复旦大学经济学院

通讯地址：上海市杨浦区国权路 600 号

邮政编码：200433

联系电话：021-65642334

电子邮件：cs@fudan.edu.cn

作者简介：陈硕，副教授，复旦大学经济学院。2011年毕业于香港科技大学社会科学部，获社会科学博士学位。主要研究兴趣集中在中国的发展经济学、政治经济学及经济史，其论文见《American Political Science Review》、《经济研究》、《管理世界》、《世界经济》、《经济学季刊》、《统计研究》等学术刊物。

实证研究中的语言逻辑与数理逻辑

From Intuitive to Empirical: Some Thoughts on KKV1994

【摘要】作为《Designing Social Inquiry: Scientific Inference in Qualitative Research》的简体版译者，我在本文中试图对学者在实证研究中经常遇到的语言逻辑和数理逻辑这两个概念进行初步的讨论。这两个概念并不是直接来自于原书内容，而是我在学习及翻译过程中获得的最重要启发。语言逻辑往往采用文字表达方式阐述因果关系，而数理逻辑则依赖于更加严谨的公式推导或研究设计。语言逻辑同时为没有受过社会科学训练的普通公众及专业学者所使用，前者更多依赖于常识及经验，在多数情况下无法准确判断因果关系；后者则构成日常学术交流的主要组成部分。对于后者使用的语言逻辑，本文认为通过因果思维或研究设计将其转化为数理逻辑是决定实证研究者研究水平的重要因素。

Abstract: As the translator of *Designing Social Inquiry: Scientific Inference in Qualitative Research*, in this paper, I attempt to clarify two different concepts – intuitive expression and mathematical thinking. Intuitive expression refers to the principle of logic and logical thoughts embedded in the natural language deriving from the common sense and personal experiences, while mathematical thinking mainly resorts to deliberate formula derivation and more scientific research design. I argue that only intuitive expression cannot draw the credible causal inference. Mathematical logic, as the most common-used proofing method, is often employed in the daily academic research. The ability to convert the intuitive idea to a mathematical way is an important criterion for a qualified researcher.

【关键词】 社会科学 因果推论 语言逻辑 数理逻辑

Key Words: Social Science; Causal Inference; Intuitive Expression; Mathematical Thinking

作为《Designing Social Inquiry: Scientific Inference in Qualitative Research》的简体版译者，我非常荣幸能够在这本书出版 20 周年之际将她介绍给国内的读者。该书（以三位作者姓氏首字母及出版时间构成的简称 KKV1994 也许在学界更具知名度）由普林斯顿大学出版社在 1994 年出版。KKV1994 在出版后就成为众多欧美高校经济学、政治科学及社会学等学科本科及研究生方法课程的指定参考书，也是当今最具影响力的社会科学方法论著作。这本书是我在香港科技大学社会科学部硕士阶段的必修课《社会科学方法论》的指定教材，也是我读的第一本方法论著作，对此我感到非常幸运。就内容来说，三位作者 Gary King、Robert Keohane 和 Sidney Verba 在第一章便明确指出该书的写作目标是“向读者系统地介绍如何在社会科学领域开展研究”。而对于研究实施的两个方法：定量传统及定性传统，作者认为它们“之间的差别仅仅是风格和方法论上的不同...所依赖的逻辑都是相同的”：“即怎样提出研究问题、开展学术研究并进行有效的描述及因果推论”。不论采用何种方法，评价其研究过程是否是科学的主要标准都要看其是否满足：1) 以推论为研究目的；2) 采用公开的研究程序；3) 结论是不确定的；4) 研究的内容是关于方法的。

围绕上述这些观点，本书剩余部分由两个方向展开：第一个方向将研究过程分为描述性推论（第 2 章）及因果性推论（第 3 章）两个步骤，其中前者是研究的起点并构成后者的基础。在作者看来，这两个步骤均存在推论的过程：描述性推论涉及如何由样本推及整体，而因果推论则试图发现不同因素间真实的因果关系。本书第 2 及第 3 章提供了这两个推论的定义及推论规则；第二个方向为主题研究，作者向读者提供处理研究中经常遇到的特定问题的对策。这些问题包括：如何确定观测对象（第 4 章）、如何增加观察值数量（第 6 章）及如何避免某些特定问题（测量误差、遗漏变量、无关变量及内生性等，见第 5 章）。就如何确定观察对象来说，作者指出在大多数定性研究中随机选择样本方法具有一定局限性，在这种情况下有目的地选择样本也许更为可取。围绕有目的取样，作者提供了 5 种可供读者参考的操作方案。就研究中需要避免的问题来说，测量误差、遗漏变量、无关变量及内生性问题要么导致推论效率降低，要么导致更严重的有偏推论。这些问题在定量研究中尤为突出，需要值得读者重视。辅之以具体的例子，作者阐述这些问题造成的上述后果并提供了对策。而最后一章讨论的增加样本数量则是定量及定性学者面临的共同问题，如何在现有限制下尽量拓展样本数以提高推论效率的方法会在这一章给出。

我把这本书的中译书名定为《社会科学中的研究设计》的目的也正是为了消除定量及定性研究方法间的隔阂。不管是定量还是定性背景，我确信每位读者都可以从这本书中汲取所需的养分。对我而言，本书的学习及翻译过程带给我最重要的财富是让我具备了用因果思维的能力，特别是语言逻辑和实证逻辑之间的切换。语言逻辑主要采用文字或者口语表达出来的因果效用，主要由两个群体所使用：没有受过社会科学方法论训练过的公众依据习惯、阅历及常识判断出来的描述及因果推论以及受过专业训练的学者在学术活动时采用的交流方式。这些学术活动即包括学者日常的交流、学术研讨会上的提问与质疑，也包括学术期刊审稿意见用语；而数理逻辑则是受过训练的定量研究者利用实证工具，包括公式表达或研究设计对语言逻辑进行严谨表述及应对过程。¹这两个逻辑正是原书作者整合定性和定量研究思路的努力给我带来的直接启发。我认为语言逻辑和数理逻辑贯穿了实证学者研究生涯的全部过程。而快速地实现这两种逻辑间的转化不光需要大量的学术阅读，更需要有针对性的训练。

¹ 从更严谨的角度来说，本文使用的“语言逻辑”近似于逻辑学中的“普通逻辑”：用自然语言研究人们日常思维形式及规律的逻辑，包含较多的经验及常识成分。“数理逻辑”则基本等同逻辑学中的“数理逻辑”概念。二者构成了“形式逻辑”的主要内容。此点感谢审稿人补充。

在本文的以下部分，我将首先阐述常识在很多情况下对判断因果效用帮助不大；其次，在更多情况下，没有常识能够帮助研究者界定出因果关系。此时，对因果效用的识别，特别是大小的考察需要依赖数理逻辑。我将结合测量偏误、互为因果及遗漏变量这三个研究者经常遇到的问题加以阐述。²众所周知，这三个问题将导致内生性问题：最小二乘估计（OLS）估计结果不一致。不一致的结果是最严重的估计偏差问题：偏差并不会随着样本量趋于无穷大而消失。在这种情况下，研究者无法根据 OLS 估计系数并结合偏差方向来推测真实效应，从中得到的政策含义也将具有误导性；最后，我会指出基于因果推论思维的语言逻辑是构成学者学术生活的重要组成部分。在这种情况下，实现两种逻辑间的转化是决定实证研究者研究水平的重要因素。对于该结论，我将结合自己研究中遇到的一个具体事例加以阐述。

常识并不可靠

基于语言逻辑的常识、习惯或阅历的确能够帮助我们判断一部分关系中的因果方向，但有些情况下它们并不可靠。一个典型例子是财富和幸福的关系。常识往往认为财富能够带来幸福。实际上，二者间的关系是受到诸多个体特质的同时影响：勤劳、人格、健康状况、自控及自信等等（Diener, 2012; Luhmann *et al.*, 2012）。正是这些没有被常识所考虑的因素同时驱动着个体获得财富并维持幸福感（Diener and Fujita, 1995）。将该遗漏变量纳入考察范围之后，现有研究已经发现在满足基本需要后的财富增长带来的幸福感是微乎其微的（Roberts, 2011）。这个例子就说明常识要不就是在诉说一个根本不存在因果效应的关系要不就是极大高估了金钱对幸福的作用。

没有常识可以依靠

在其他绝大多数情况下，对特定关系是否存在真实因果效应可能完全没有常识可依靠的。测量偏误及互为因果问题就是这样的例子。语言逻辑无法帮助研究者廓清因果效应。此时，基于公式推导的数理逻辑能够展示 OLS 估计系数存在的偏差情况。

测量偏误简单说来就是由于某些原因造成变量的测量值和真实值之间存在差异。造成偏误的原因有很多，比如测量技术局限、原始资料局限、数据录入人员的不小心或受访人的故意虚报。当存在这个问题时，如何判断对因果推论的影响及寻找相应的解决手段就至关重要了。如果测量偏误随机存在的，比如数据录入人员的粗心。我们一般将这种测量误差称之为经典测量误差，它将导致该变量作用的低估。这里我们用收入对消费影响的例子加以演示。假设二者间关系的 OLS 表达式如下：

$$y_i = \beta_0 + \beta_1 x_{li}^* + u_i \quad (1)$$

其中 y_i 是个体的消费，而 x_{li}^* 是收入， u_i 是其他可能影响消费水平但是没有被模型捕获的因素，假设是随机分布的。为了简单，我们也假设收入-消费关系不存在其他影响估计结果一致性的问题，此时，

² 大部分计量经济学教科书对于这三个问题的数理推论过程都大同小异。本文的推论基于 Wooldridge (2012)，但根据实际例子作了修正。

$$e_1 = x_1 - x_1^* \quad (2)$$

其中 x_1 是收入的真实值， e_1 是真实值和录入值之间的差异，也就是测量偏误。在录入原始问卷过程中，由于数据录入人员的粗心导致一些收入观察值的录入错误。如果我们认为这种不小心的分布是随机的。那么，真实值和录入值就均独立于 e_1 ，写作：

$$Cov(x_1, e_1) = 0 \quad (3)$$

$$Cov(x_1^*, e_1) = 0 \quad (4)$$

将 (2) 带入 (1) 得到：

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1) \quad (5)$$

基于 (5)，计算 x_1 和 e_1 的协方差：

$$Cov(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_{e_1}^2 = \sigma_{e_1}^2$$

那么：

$$Cov(x_1, u - \beta_1 e_1) = -\beta_1 Cov(x_1, e_1) = -\beta_1 \sigma_{e_1}^2$$

由此计算出收入对消费作用 β_1 的 OLS 估计结果概率极限为：

$$\begin{aligned} p \lim(\hat{\beta}_1) &= \beta_1 + \frac{Cov(x_1, u - \beta_1 e_1)}{Var(x_1)} \\ &= \beta_1 - \frac{\beta_1 \sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} = \beta_1 \left(1 - \frac{\sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right) \\ &= \beta_1 \left(\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right) \end{aligned}$$

很明显， $\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}$ 介于 0 和 1 之间。由此可得在经典测量误差情况下，OLS 估计结果将小于真实作用大小。换句话说，OLS 估计结果是真实结果的下限。

并不是在所有情况下，我们都能假定测量误差是随机分布的。误差的不随机性意味着(4)式不再成立，进而导致 $Cov(x_1, e_1) \neq \sigma_{e_1}^2$ ，最终 $p \lim(\hat{\beta}_1) \neq \beta_1 \left(\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} \right)$ 。此时，研究者无

法判断 OLS 估计值相对于真实值的偏离方向。如何应对则取决于具体研究议题。Ashenfelter 和 Krueger 于 1994 年发表在《American Economic Review》上的著名论文《Estimates of the Economic Return to Schooling from a New Sample of Twins》采用工具变量方法（Instrumental Variable Approach）处理该问题。作者在该文中试图估计教育对个体收入的作用。由于担心受访者自己汇报的教育年限可能存在测量误差且该误差并不遵从随机分布：教育程度较低的受访者倾向高报自己的受教育水平，而教育程度较高的受访者则无此动机。作者便采用受访者双胞胎兄弟姐妹提供的受访者教育程度作为受访者真实教育程度的工具变量。在作者看来，这两个变量之间的相关系数较高，同时更加重要的是受访者兄弟姐妹提供的教育年限对受访者收入没有直接影响。³另外一个可供参考的应对非经典测量偏误问题方法是利用子样本重复回归。我和伦敦政治经济学院的马德斌教授合作的一篇工作论文《War Making: the Formation of Big Oriental Despotism, 750 BC-1911 AD》便采用了这一方法。该文试图利用-750 年至 1911 年的历史数据实证检验战争对中国国家形成的影响。战争的原始数据来自中国的历史文献：春秋战国时期的数据基本来自于《左传》，后期朝代的战争数据来源较多包括历代《实录》、《二十四史》、《史记》、《资治通鉴》及地方官员奏折等。由于数据来源的局限，和明清时期的战争数据相比，汉之前的数据可信性较差也无其他数据源可以相互补充验证。当然，研究者可以通过控制朝代固定效用，甚至是皇帝固定效用来控制那些不随着时间变化且影响记录质量的因素。另外一个更为简便的可行方案就是只采用数据质量较高的明清时期子样本重新检验战争和国家形成的关系。然后利用一些统计技术比如 Hausman 检验来比较基于子样本获得的系数和全样本系数间是否存在系统性差异。如果存在差异，研究者至少需要警惕测量偏误带来的风险并谨慎解释基于全样本的估计系数。

另外一个无法依赖于常识判断估计结果偏差方向的例子是同时性问题。同时性指的是解释变量和被解释变量互相影响，也被称之为互为因果问题。该问题的一个典型例子是警察或者司法支出与犯罪之间的关系：一方面，警察的增加可以增加威慑、提高破案时间进而减少犯罪（Levitt, 1997; Levitt, 2002）。此时，二者之间是负向相关；另一方面，一个地区犯罪率的增加也会导致政府招募更多的警察及提高司法支出。此时，二者之间呈现出正向关系。虽然从逻辑上两个因果方向都能自圆其说，但其背后的政策取向截然不同。如果是前者，降低犯罪的有效应对措施就是增加警察及司法支出；后者则意味着警察和司法支出的增加成了犯罪增加的结果，而非原因。造成这个地区犯罪猖獗的根源可能是不平等、社会福利、受教育水平及外来人口等深层次因素。此时政府一味提高司法支出不仅不会导致犯罪的下降而且造成了公共财政资金利用的低效。

那么互为因果问题的存在为什么会致 OLS 估计结果不一致呢？回答这个问题同样需要借助数理逻辑。司法支出和犯罪之间相互影响关系可以写作如下 OLS 方程组：

³ 另外一个用工具变量处理核心解释变量测量偏误的研究见陈硕和陈婷（2014）。

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \quad (6)$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

假设 y_1 是犯罪率， y_2 是司法支出， z_1 是影响犯罪率的其他因素而 z_2 是影响司法支出的其他因素。 u_1 和 u_2 均符合 OLS 的基本假设。此时有：

$$y_2 = \alpha_2 (\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2 \quad (7)$$

(7) 可进一步写成：

$$(1 - \alpha_2 \alpha_1) y_2 = \alpha_2 \beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2$$

为了表述的方便，该式可整理成：

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2 \quad (8)$$

其中：

$$\pi_{21} = \frac{\alpha_2 \beta_1}{1 - \alpha_2 \alpha_1}, \pi_{22} = \frac{\beta_2}{1 - \alpha_2 \alpha_1}, v_2 = \frac{\alpha_2 u_1 + u_2}{1 - \alpha_2 \alpha_1}$$

此时，基于 (8) 计算 y_2 和 u_1 的协方差：

$$\begin{aligned} Cov(y_2, u_1) &= Cov(y_2, v_2) = \left(\frac{\alpha_2}{1 - \alpha_2 \alpha_1} \right) E(u_1^2) \\ &= \left(\frac{\alpha_2}{1 - \alpha_2 \alpha_1} \right) \sigma_1^2 \end{aligned}$$

我们知道一致地估计司法支出对犯罪的作用需要的假设是 y_2 和 u_1 相互独立，即 $Cov(y_2, u_1) = 0$ （可参见 Wooldridge, 2012）。但除非 $\alpha_2 = 0$ ，二者之间无法实现独立。而 $\alpha_2 = 0$ 恰恰意味着犯罪并不影响司法支出：互为因果不存在。

从研究设计上，互为因果问题是最为棘手的内生性问题。目前较为流行的处理方式也是工具变量方法。我在 2012 年发表在《经济学（季刊）》上的论文《转型期中国的犯罪治理政策：堵还是疏？》中试图实证检验中国司法支出的效率。由于上文提到的司法支出和犯罪之间相互影响，我在文中使用省政法委书记党龄作为司法支出的工具：政治资历和获得的资源成正比同时犯罪率不会反响影响其政治资历。基于工具变量方法，该文发现司法支出对犯罪率没有显著影响。相比之下，不断增长的犯罪率正很大程度上是转型期多种社会经济特征所致。

4

此外，在一些情况下虽然常识可以判断因果方向但并不能告诉方向的大小。科学的严谨性促使研究者在方向的基础上进一步估计出作用大小，后者在社会科学中尤为重要。实现对大小的估计只能依赖于数理逻辑。以下将以教育回报为例阐述该观点。每个人都知道教育对个体收入的回报为正，但经验无法回答每增加一年教育时限到底会增加百分之多少的收入。可能很多人不理解社会科学家对小数点后几位数字的大小争论不休的意义何在。设想一下如果事关数以千亿元的教育支出，小数点后面数字就不再那么无关紧要的。对教育回报大小的估计是劳动经济学数十年来关注的经典研究问题之一，该学科为了回答该问题进行的研究设计及方法论上的创新也极大地推动了其他社会科学领域的发展。

对教育回报作用识别遇到的最棘手问题是遗漏变量，此时遗漏的这个变量便是能力。一般来说，高能力个体同时具有较长的教育年限及较高的收入。那么观察到的教育作用就在一定程度上提取了能力的作用，如果忽视该变量将导致对教育程度作用的高估。显而易见，基于高估的系数制定的教育支出预算必然导致巨大的公共财政浪费。在能力变量的干扰下，即使没有社会科学训练的人也不难理解为何教育程度会被高估。但语言逻辑表达出来的高估只有通过数理逻辑展示出来才能在更一般的意义上理解遗漏变量问题对因果推论的影响。我们假设教育和收入关系的 OLS 表达式如下：

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \quad (9)$$

其中 y_i 是收入，比如为小时工资， x_{2i} 是教育年限， x_{3i} 则是上文提到的能力。假设 u_i 符合 OLS 一般假设。在实际研究中能力 x_{3i} 无法观测或可以观察却无法准确地测量因此没有包含在方程之中，也就是说 $\beta_3 x_{3i} + u = u_i^*$ 。

此时的 (9) 就变成了：

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i^* \quad (10)$$

对 (10) 进行 OLS 估计，获得 β_2 的 OLS 估计值是：

⁴ 另外一个同样存在反向因果的相关研究问题是惩罚与犯罪，见陈硕和章元（2014）及陈硕（2014）。

$$\hat{\beta}_{2,ols} = \frac{\sum(x_{2i} - \bar{x}_2)(y_i - \bar{y})}{\sum(x_{2i} - \bar{x}_2)^2}$$

其中， $y_i - \bar{y} = \beta_2(x_{2i} - \bar{x}_2) + \beta_3(x_{3i} - \bar{x}_3) + u_i - \bar{u}$

那么

$$\begin{aligned}\hat{\beta}_{2,ols} &= \beta_2 + \frac{\beta_3 \sum(x_{2i} - \bar{x}_2)(x_{3i} - \bar{x}_3) / n}{\sum(x_{2i} - \bar{x}_2)^2 / n} + \frac{\sum(x_{2i} - \bar{x}_2)(u_i - \bar{u}) / n}{\sum(x_{2i} - \bar{x}_2)^2 / n} \\ &= \beta_2 + \beta_3 \frac{\text{sample cov}(x_{2i}, x_{3i})}{\text{sample var}(x_{2i})} + \frac{\text{sample cov}(x_{2i}, u_i)}{\text{sample var}(x_{2i})}\end{aligned}$$

当 $n \rightarrow \infty$ 的时候， $\hat{\beta}_{2,ols}$ 趋近为 $\beta_2 + \beta_3 \frac{\text{cov}(x_{2i}, x_{3i})}{\text{var}(x_{2i})} + \frac{\text{cov}(x_{2i}, u_i)}{\text{var}(x_{2i})}$

此时， $p \lim \hat{\beta}_{2,ols} = \beta_2 + \beta_3 \cdot b_{32}$ ， b_{32} 是 x_{2i} 对 x_{3i} OLS 回归获得的系数的概率极限。

从上可知除非 $\beta_3 = 0$ 或 $b_{32} = 0$ ，否则 $\hat{\beta}_{2,ols} \neq \beta_2$ 。而只有当 $\text{cov}(x_{2i}, u_i^*) = 0$ 时， b_{32} 才等于 0。由于能力的作用 β_3 以及 b_{32} 都是大于 0 的：高能力的收入和教育程度都高。这种情况下， $p \lim \hat{\beta}_{2,ols}$ 大于 β_2 ：OLS 高估了教育回报的作用。

综上所述，对不可靠或者无法依靠的常识，研究者必须了解其背后的数理逻辑。当然，并不是所有场合都需要复述出上述公式的详尽推导过程，但研究者至少需要知道无法依靠或估计存在偏差在数理逻辑中到底体现在哪个地方。高素质的研究者应当尽量缩短上述转化所需要的时间并提出解决方案，前者需要对上述推导了然于心，后者则需要结合大量文献阅读及有针对性训练。

语言逻辑和数理逻辑间的转换是决定实证研究者研究水平的重要因素

对于研究者来说，其学术和研究生涯遇到大多数反馈都是基于语言逻辑。比如，在各种学术讨论上、与同行交谈甚至在审稿人意见中，对方一般用语言的方式陈述作者研究工作中存在的问题：“你的结果高估了真实作用大小”、“你的解释变量存在内生性问题”、“你的因果关系可能只是相关”、“样本也有自选择问题”等等。这样的语言逻辑和上文提到的常识不同：使用者为受过专业训练的社会科学研究人员而非一般公众。在这种情况下，如何将这

阐述成严谨的公式推导对于研究者回应问题、重新研究设计以增加稳健型讨论就非常重要了。我和香港科技大学社会科学部 James Kung 教授合作的工作论文《Of Maize and Men: The Effect of a New World Crop on Population and Economic Growth in China》在审稿过程中便遇到这样情况。长期以来，马尔萨斯型社会中的技术进步所造成的经济和社会影响一直为经济史学家和宏观经济学家所关注：理论模型预期在马尔萨斯时代，技术进步仅会带来人口增加但不会改善长期人均生活水平。这篇论文试图通过实证检验北美作物玉米在中国引种和拓展对明清时期人口和经济发展的作用以检验上述理论。玉米由于其抗旱性和对地形的低要求，被广泛种植于清代内陆的新开发地区并显著提高了中国的粮食产量：1900 年中国粮食生产面积比 1380 年增加 3.3 倍，对同时期粮食增加的贡献率超过 55% (Wang, 1973; Perkins, 1969)。学者因此将北美作物在中国的引种和传播比作是宋代引种占城稻之后的第二次农业技术革命 (Ho, 1959; Lan, 2002)。实证考察玉米作用需要该作物在中国引种的系统数据。当我在进行这个研究的时候没有任何一本历史著作系统搜集过这些数据。历史学家仅仅提供了大概引种趋势：玉米在 16 世纪中叶被引种到中国，经过 3 个多世纪的拓植，在 20 世纪初期中国的所有地区均种植了玉米 (Ho, 1959)。虽然这些研究富有先导性和启发意义，我们需要的是玉米在中国每个地区的精确引种时间，这无疑是一次劳动密集型工作。我在博士阶段用了整整两年的时间从明清时期出版的全部 1119 本府志中诸册查找并最终建立了玉米引种的数据库 (图 1)。



1580



1600



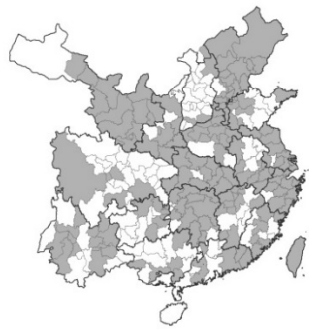
1650



1700



1750



1800



1850



1900

图 1：玉米在中国的引种

这篇工作论文在第一轮审稿中一共获得了 4 个审稿人的 142 条审稿意见，其中绝大多数都是语言逻辑：他们仅仅指出初稿中他们认为有问题的地方。如何用实证方式加以处理则是作者自己的事情。论文最终能否被期刊所录用就取决于能否令审稿人满意地处理他们的质疑。以下用其中一条意见为例阐述这两种逻辑之间转化的重要性。

图 1 展示出的这种渐进时空变化使得研究者可以采用标准的双重差分方法（Difference in Differences）来检验其作用，其作用估计结果见图 2：

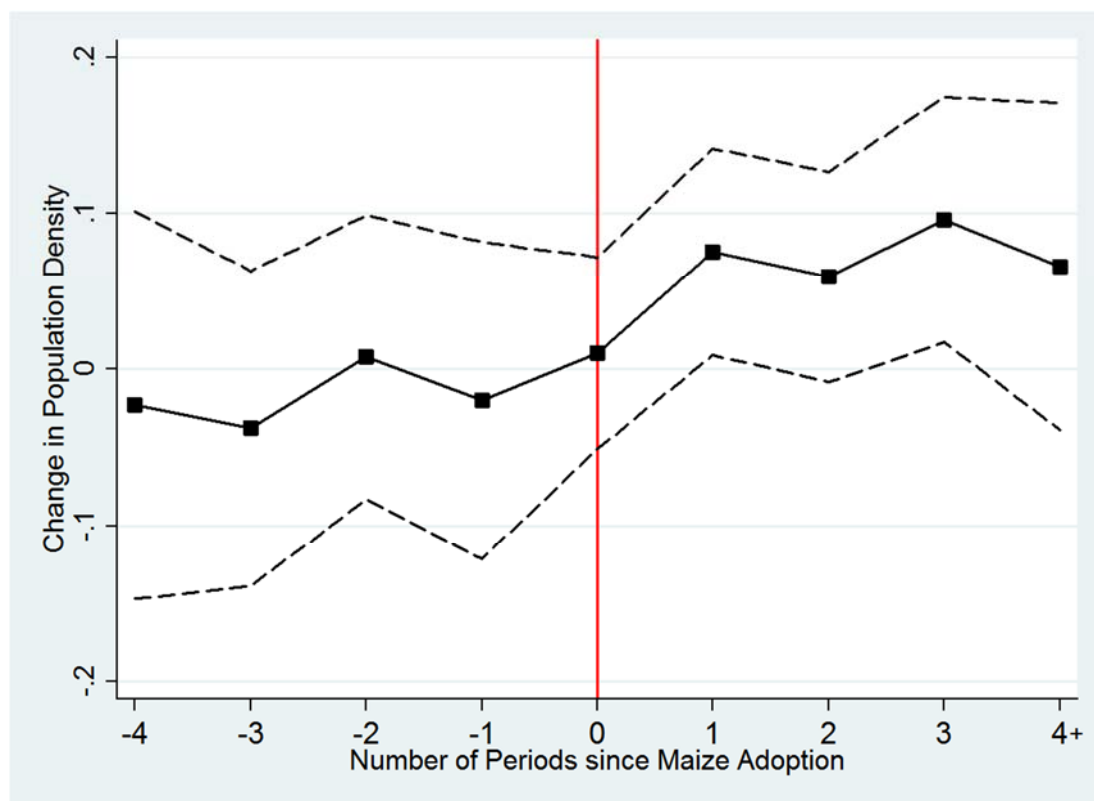
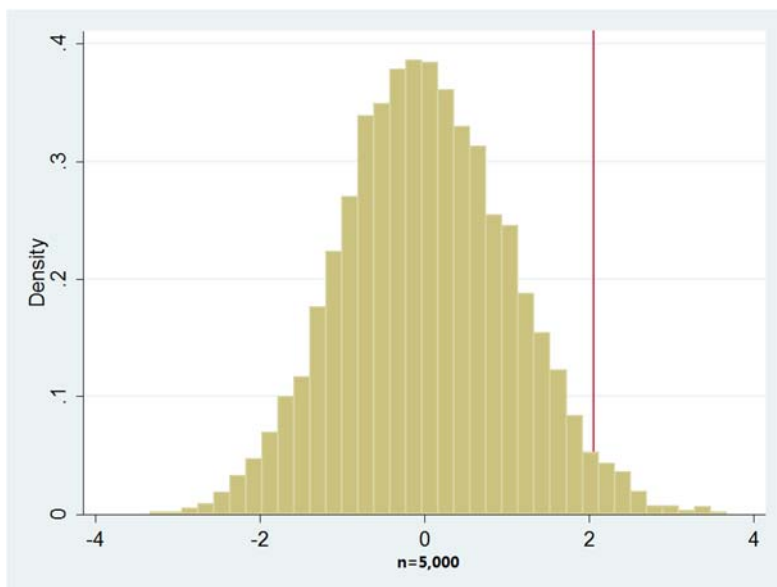
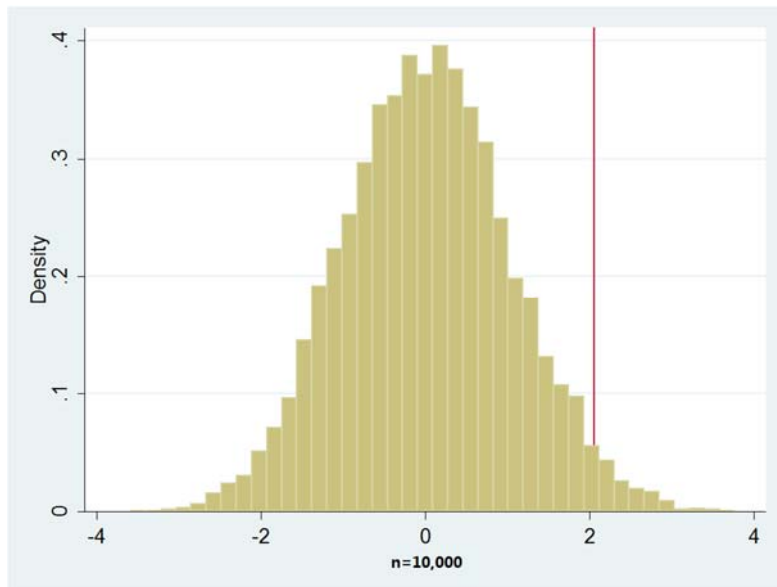


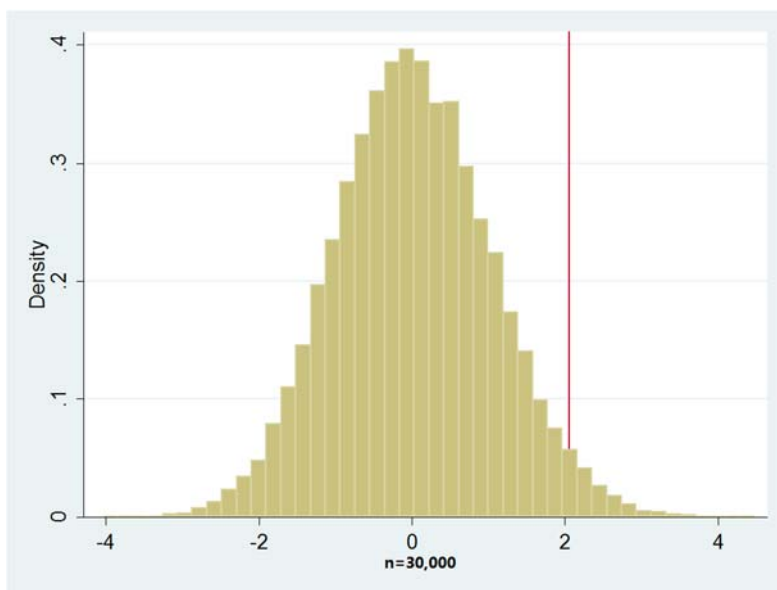
图 2：玉米引种前后的人口变化

图 2 的横坐标表示距离该地区玉米引种的时期数，纵坐标则表示人口密度在引种和没引种地区之间的差异。图中的实线表示人口密度在玉米引种前后的变化，两条虚线分别为估计系数对应的上下 95% 置信区间。从该图中，我们可以看到在玉米引种之前，两个地区之间没有显著差异。但在引种之后，引种地区的人口密度开始超过没有引种玉米的地区。在工作论文里面，该图出现在正式回归之前作为玉米促进人口增长的描述性证据。其中一个审稿人对此提出质疑：他注意到在玉米引种前的 3 至 2 期及 1 期到 0 期，引种地区和非引种地区人口的差别在趋势上同样存在上升趋势。在这种情况下，审稿人认为我们无法排除从 0 到 1 期的人口增长可能是偶然因素所致而非是玉米作用。

在我的理解中，如果 0 至 1 期的正斜率来自于随机效果的话，那么把一个地区的玉米真实引种时间随机赋给另外一个地区然后重新回归依然得到相似效果，如果真是这样的话，我们就无法再坚持 0 到 1 期的正斜率是玉米的作用了。为了实现概率上的效果，我们需要重复这样的赋值很多次。图 3 展示该项虚拟工作获得的玉米引种估计系数的 t 值，三个图依次是进行 5000 次、10000 次及 30000 次随机赋值的结果。从中可以清楚的发现，在

这种重复实验中，获得 t 值呈现正态分布：绝大多数都介于 -2 和 2 之间。这意味着随机复制获得的估计系数绝大多数都不显著。虽然在概率上仍旧可以坚持基于真实赋值获得的 0 到 1 期正效应还是偶然的。但这种偶然性太低了，低到足以使得我们认为这是不可能的：该时段的正效用的确由于玉米引种所致。在该论文的第三次审稿中，上述修改意见最终得到了审稿人的认可。





注：三条红线表示玉米在实际回归中获得的估计系数， t 值的绝对值大于这三个系数的概率分别是 4.64%、4.24% 及 4.55%。

图 3：玉米作用的虚拟赋值实验

严谨的因果思维能力是高度训练后养成的自觉。了解文字逻辑对应的数理逻辑需要研究者接受刻苦的方法论训练。因此，在研究开始之前，系统学习线性代数、概率论及计量经济学的知识是必须的。但实现这两种逻辑之间的转换对研究者的要求则更高：坚实的计量经济学基础之外仍需要持之以恒且大量地学术阅读，以此培养实际应对能力。从这个意义上，求知欲和进取心只是第一步，除此之外还需要一个好的老师领你入门。我想说的是，KKV1994 正充当这种教练的角色。对于初学者来说，这本书提供了训练所必需的知识，同时又维持在一个恰当的难度水平之上。作者们主要借助浅显易懂的文字逻辑而非抽象复杂的数理逻辑告诉读者在研究设计中会遇到的问题，造成的后果及相应的对策。在维持读者信心的同时，循序渐进地揭示出描述性推论及因果推论的相关知识。KKV1994 的风格使我从没感到方法论很枯燥。恰恰相反，它是我认为最有意思的一门学问。同时，对于学者来说，这也是一本常读常新的著作。工作以后，我依然经常抽空翻翻它。书里面的内容不断提醒自己在研究过程中要避免作者们提到的那些问题。这种互动，正如作者在中文版序言中所写：“本书的信息可以促进读者重新审视自己的研究，同时在有了实际研究经验之后重读本书还能获得新的感悟。”

参考文献

- 陈硕(2014a). 社会科学中的研究设计. 上海：上海人民出版社.
- 陈硕(2014b). 转型期中国的犯罪问题的经济解释：行为分析及治理政策. 北京：中国社会科学出版社.
- 陈硕(2012). 转型期中国的犯罪治理：堵还是疏. 经济学（季刊），2.
- 陈硕、陈婷(2014). 增长约束下的环境治理：气体排放与公共健康. 经济研究，8.
- 陈硕、章元(2014). 治乱无需重典：转型期中国刑事政策效果分析. 经济学（季刊）. 4.
- Ashenfelter, O., & Krueger, A. (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review*, 1157–73.
- Chen Shuo & Kung James (2015). Of Maize and Men: The Effect of a New World Crop on Population and Economic

- Growth in China. Fudan University, Department of Economics, Working Paper.
- Chen Shuo & Ma, Debin (2015). War Making: the Formation of Big Oriental Despotism, 750 BC-1911 AD. Fudan University, Department of Economics, Working Paper.
- Diener, D. (2012). New Findings and Future Directions for Subjective Well-being Research. *American Psychologist*, 67(8): 590–97.
- Diener, E., & Fujita, F. (1995). Resources, Personal Strivings, and Subjective Well-being: a Nomothetic and Idiographic Approach. *Journal of personality and social psychology*, 68(5): 926.
- Ho, Ping-ti (1959). *Studies on the Population of China, 1368–1953*. Cambridge: Harvard University Press.
- Lan, Yong (2002). *Historical Geography of China (Zhongguo Lishi Dilixue)*. Beijing: Higher Education Press.
- Levitt, S.D. (1997). Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime, *American Economic Review*, 87(3): 270–290.
- Levitt, S.D. (2002). Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply, *American Economic Review*, 92(4): 1244–50.
- Luhmann, Maike; Hofmann, Wilhelm; Eid, Michael; Lucas, Richard E. (2012). Subjective Well-being and Adaptation to Life Events: A Meta-analysis. *Journal of Personality and Social Psychology*, 102(3): 592–615.
- Perkins, H. Dwight (1969). *Agricultural Development in China 1368–1968*. Edinburgh: Edinburgh University Press.
- Roberts, J. A. (2011). *Shiny objects: Why We Spend Money We Don't Have in Search of Happiness We can't Buy*. New York: Harper Collins Publishers.
- Wang Yeh-chien (1973). *Land Taxation in Imperial China, 1750–1911*. Cambridge: Harvard University Press.
- Wooldridge M. Jeffrey (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning.