

当代经济学系列丛书

Contemporary Economics Series

主编 陈昕



基本无害的计量经济学 实证研究者指南

当代经济学
教学参考书系

[美] 乔舒亚·安格里斯特 著
约恩-斯特芬·皮施克

郎金焕 李井奎 译



格致出版社
上海三联书店
上海人民出版社

当代经济学系列丛书

Contemporary Economics Series

主编 陈昕

基本无害的计量经济学 实证研究者指南

[美] 乔舒亚·安格里斯特 著
约恩-斯特芬·皮施克

郎金焕 李井奎 译

当代经济学参考书系



格致出版

格致出版社

上海三联书店

上海人民出版社

图书在版编目(CIP)数据

基本无害的计量经济学:实证研究者指南/(美)
安格里斯特,(美)皮施克著;郎金焕,李井奎译. —上
海:格致出版社:上海人民出版社,2012
(当代经济学系列丛书/陈昕主编.当代经济学教
学参考书系)

ISBN 978-7-5432-2058-4

I. ①基… II. ①安… ②皮… ③郎… ④李…
III. ①计量经济学-高等学校-教材 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2012)第 013012 号

责任编辑 钱 敏
装帧设计 敬人设计工作室
吕敬人

[美]乔舒亚·安格里斯特 著
约恩-斯特芬·皮施克
郎金焕 李井奎 译

基本无害的计量经济学:实证研究者指南

格致出版社·上海三联书店·上海人民出版社
(200001 上海福建中路 193 号 24 层 www.ewen.cc)



编辑部热线 021-63914988
市场部热线 021-63914081
www.hibooks.cn

世纪出版集团发行中心发行
苏州望电印刷有限公司印刷

2012 年 4 月第 1 版

2012 年 4 月第 1 次印刷

开本:787×1092 1/16

印张:17 插页:5 字数:391,000

出版前言

为了全面地、系统地反映当代经济学的全貌及其进程,总结与挖掘当代经济学已有的和潜在的成果,展示当代经济学新的发展方向,我们决定出版“当代经济学系列丛书”。

“当代经济学系列丛书”是大型的、高层次的、综合性的经济学术理论丛书。它包括三个子系列:(1)当代经济学文库;(2)当代经济学译库;(3)当代经济学教学参考书系。该丛书在学科领域方面,不仅着眼于各传统经济学科的新成果,更注重经济学前沿学科、边缘学科和综合学科的新成就;在选题的采择上,广泛联系海内外学者,努力开掘学术功力深厚、思想新颖独到、作品水平拔尖的“高、新、尖”著作。“文库”力求达到中国经济学界当前的最高水平;“译库”翻译当代经济学的名人名著;“教学参考书系”则主要出版国外著名高等院校的通用教材。

本丛书致力于推动中国经济学的现代化和国际标准化,力图在一个不太长的时期内,从研究范围、研究内容、研究方法、分析技术等方面逐步完成中国经济学从传统向现代的转轨。我们渴望经济学家们支持我们的追求,向这套丛书提供高质量的标准经济学著作,进而为提高中国经济学的水平,使之立足于世界经济之林而共同努力。

我们和经济学家一起瞻望着中国经济学的未来。

前 言

计量经济学的疆界正在不断扩张。作为这种扩张的结果,其方法和实践也有了长足发展,但即使是那些精于数据处理的个中老手,也会对如今如此繁多的计量方法感到困惑。幸运的是,并非所有方法都同样有用、同等重要。那些过于新奇的方法本来没必要如此复杂,而且还可能是有害的。从积极的方面讲,虽然对计量经济学基本工具的解释日趋精奥深微,但应用计量经济学(Applied Econometrics)的核心内容却保持着大体稳定。本书为实证研究者把握计量经济学的精义提供了一个向导,这些计量经济学的精义也就是我们所指的基本无害的计量经济学(Mostly Harmless Econometrics)。

在应用计量经济学家的工具箱中,最重要的几件工具可以列举如下:

- (1) 为了将可能掩盖因果关系的变量控制起来,而设计的回归模型(Regression Model);
- (2) 用于分析真实实验以及自然实验的工具变量方法(Instrumental Variables Method);
- (3) 在重复观察中用以处理不可观察的缺失变量的双重差分方法(Difference-in-Difference Strategies)。

对上面这些基本技巧的创造性使用要求读者对统计推断的作用机理有坚实的概念基础和良好的理解。应用计量经济学在这两方面的特点将会在本书中得到体现。

我们对计量经济学中那些内容重要的看法来自我们作为实证研究者的研究经验,而且特别来自我们的教学实践和指导经济学博士研究生的工作。正是

在与这些同学的思维交流中，我们完成了本书的写作。与此同时，我们还希望这本书能够吸引其他领域中正在苦苦探索如何选择计量方法、如何解释研究结果的研究者们。应用计量经济学所考虑的问题和其他社会科学或者流行病学所考虑的问题并无本质上的区别。任何希望运用数据指导公共政策或者推动公共卫生事业的人都要理解并使用统计结果。任何希望从数据中得到有用推断的人都可称为应用计量经济学家。

许多计量经济学方面的教科书都对研究方法提供一些指导，因此本书和其他广泛使用的教科书存在一些内容上的重叠。但这本书在多个方面有别于传统的计量经济学教科书。首先，我们认为使用数据回答特定因果关系的经验研究最有价值，这类似于在医学研究中经常出现的随机临床实验。我们研究所有问题的方法都体现这个观点。在缺乏真实实验时，我们寻找经过良好控制的对照组，或者说自然的“准实验(quasi-experiment)^①”。当然，一些准实验研究设计要比其他一些方法更有说服力，但是在这些例子中计量经济学使用的方法几乎都很简单。因此，相比于其他教科书中对计量方法的处理，这本书对相应主题的讨论显得更短小更集中。我们主要对在自己的研究中读到和使用到的概念和简单的统计技巧进行强调，并与多个实证研究案例结合来解释这些观点和技巧。尽管我们对计量经济学中什么是重要的观点并未在应用经济学家中得到一致认同，但无可争议的事实是实验和准实验研究方法逐渐居于应用经济学中最具影响力的那些研究的核心。

我们要指出的第二个不同是本书在一定程度上忽略严格性。大多数计量经济学教科书都对计量模型进行严格处理。特别的，这些书对诸如线性性和同方差性等大家认为经典模型中普遍会被违背的假设进行大量讨论。虽然在行文中也会提及这些问题，但我们采取一种更加宽容和不那么迂腐的态度。能够支持我们以上述态度进行讨论的原因乃是：我们可以对计量经济学中得到广泛使用的估计值作出一个简单的解释，这一解释与模型本身并无太大关系。如果我们得到的估计值不是我们想要的那个，那么一定是做这项研究的计量经济学家错了，而不是计量经济学错了。一个典型的例子就是线性回归，它为我们提供了关于条件期望函数^②的有用信息，而不论条件期望函数究竟是什么形状。同样的，工具变量方法可以估计出经过良好定义的总体的平均因果效应，即使这个工具变量无法影响所有个体。许多应用研究者往往从直觉上理解基本计量工具在概念上的严格性，因此隐藏在严格性背后的大部分理论将不会在本书中出现。本书在处理推断问题上也有所不同，我们并不过多地考虑渐进有效性，而是用大多数篇幅考虑实际中不易处理的有限样本问题。

① 准实验是指对控制组或者所研究因素几乎无法施加控制的实验。与一般的随机实验最大的区别在于准实验无法将个体随机分配，因此观察得到的结果可能和个体的某些不可观察因素有关，也就是说，我们不能保证得到结果是将观测个体混匀后的平均值。——译者注

② 这里的条件均值函数就是一般意义上所指的总体回归函数。在线性回归中，总体回归函数被假设为关于参数线性的。

本书的预修要求是掌握概率论和统计学的基本知识。我们特别希望读者熟悉统计推断的基本概念,比如 t -统计量和标准误(standard error)。对数学期望等概率论知识的熟悉也会有所帮助,但是之外的数学知识并不要求。虽然书中对部分重要结论进行证明,但是技术性的细节并不繁难。与很多计量经济学高级教材不同,本书仅仅少量地使用线性代数。因此,我们提供的这本指南应该比与之竞争的其他书籍更易阅读。最后,我们从 Douglas Adams 的系列轻松小说中持续获得灵感,在这种心境的引导下,我们的指南可能会偶尔地缺乏一点精确性,但是要比流行于市面上的多个百科全书式的大部头(Encyclopedia Galactica Econometrica)计量经济学教科书便宜。这里还要感谢普林斯顿大学出版社同意出版我们的这本指南。

致 谢

在这本书写作过程中,我们从很多朋友和同事的意见中受益良多。感谢 Alberto Abadie、Patrick Arni、David Autor、Amitabh Chandra、Monica Chen、Victor Chernozhukov、John DiNardo、Peter Dolton、Joe Doyle、Jerry Hausman、Andrea Ichino、Guido Imbens、Adriana Kugler、Rafael Lalive、Alan Manning、Whitney Newey、Derek Neal、Barbara Petrongolo、James Robinson、Gary Solon、Tavneet Suri、Jeff Wooldridge 以及 Jean-Philippe Wullrich,他们在本书构思和写作的不同阶段都给予了反馈,当然,作者文责自负。同样的感谢还要送给我们在伦敦经济学院和麻省理工学院的学生们,他们使用了本书的最初版本并帮助我们认清楚哪些部分是重要的。我们要特别感谢技巧高超的助教 Bruno Ferman、Brigham Frandsen、Cynthia Kinnan 以及 Chris Smith。我们感激绘图师 Karen Norberg 的无私奉献,他绘制了每章开始的那些图片并且在大大小小很多事情上给予反馈。我们还要感谢普林斯顿大学出版社的编辑 Tim Sullivan 和 Seth Ditchik、本书编辑 Marjorie Pannell 以及制作编辑 Leslie Grundfest 的热情帮助。最后,感谢我们的妻子给予的爱和支持。她们比任何人都了解做一个实证研究者伴侣的滋味。

本书结构

我们从两个作为引言的章节开始。第1章描述了对之后章节可能很有用的研究设计步骤。第2章讨论了在医学研究中用到的随机实验,这个实验为我们最感兴趣的问题提供了一个理想的基准。在引言章节之后,本书第二部分共有三章,分别讨论了回归、工具变量和双重差分法的核心内容。这三章内容既强调估计值的一般性质(比如回归总是可以近似条件期望函数等),也强调了对估计值赋予因果解释所需的假设(比如条件独立假设、工具变量“就像”随机分配、相似世界等)。在本书第三部分我们转入扩展。其中第6章考察对非连续实验的回归分析,我们既可将该部分内容看作回归—控制这种研究策略的变体,也可将其看作是一类工具变量估计法。在第7章我们讨论了用分位数回归来估计我们关心的变量对被解释变量分布的影响。最后一章则针对的是统计推断问题,我们在之前章节中对渐进性质进行考察时省略了这一部分内容。本书的一些章节里包含了更具技巧性或者专门性的小节,可以在不影响掌握本书主旨的前提下省略,这部分小节都用星号(*)标出。

目 录

001	出版前言
001	前言
001	致谢
001	本书结构

第一部分 导 论

003	1 关于“问题”的问题
008	2 理想的实验
008	2.1 选择性偏误
011	2.2 用随机分配解决选择性偏误
016	2.3 对实验的回归分析

第二部分 核 心

021	3 让回归变得有意义
022	3.1 回归的基本原理
038	3.2 回归与因果关系
049	3.3 异质性与非线性
065	3.4 回归的细节
078	3.5 附录:对加权平均导函数求导

079	4 实践中的工具变量：得到你想要的
080	4.1 工具变量与因果关系
097	4.2 两阶段最小二乘的渐进推断
103	4.3 双样本工具变量和剖分样本工具变量*
105	4.4 工具变量与异质性潜在结果
122	4.5 对局部平均处理效应的推广
133	4.6 工具变量的细节
153	4.7 附录
155	5 相似世界：固定效应、双重差分和面板数据
155	5.1 个体固定效应
159	5.2 双重差分：事前与事后，处理和控制
170	5.3 固定效应与滞后被解释变量
173	5.4 附录：对固定效应模型和滞后被解释变量模型的进一步讨论

第三部分 拓展

177	6 更进一步：断点回归设计
177	6.1 清晰断点回归
183	6.2 作为一种工具变量法的模糊断点回归
190	7 分位数回归
191	7.1 分位数回归模型
200	7.2 对分位数处理效应的工具变量估计
207	8 非标准的标准误问题
208	8.1 在估计稳健标准误时存在的偏误*
218	8.2 面板数据中的聚类问题和序列相关问题
228	8.3 附录：对简单 Moulton 因子的计算
230	最后的几句话
231	术语表及名词缩写
234	参考文献
257	译后记

第一部分 导 论

关于“问题”的问题

电脑说：“我非常仔细地检查过了，它确实是我们需要的那个答案。但是，恕我直言，我认为问题的问题在于您还没有真正理解问题所在。”

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

本章简短讨论一个成功的研究项目所必备的基础。就像《圣经故事》里的《出埃及记》一样，一项研究计划可以围绕着四个问题展开。我们称它们为常见问题 (frequently asked questions, 简称 FAQs), 因为它们的确需要被研究者不断质问才行。这些问题分别是——研究对象间的关系 (relationship of interest)、理想条件下的实验 (ideal experiment)、识别策略 (identification strategy) 以及推断模式 (mode of inference)。

首先我们应该问的是：我们关注的研究对象间的因果关系是什么？尽管纯粹描述性研究也起着重要作用，但是我们相信，社会科学中最有趣的研究是关于因与果的，例如第 2 章和第 6 章所讨论的班级规模对学生分数影响的研究。因果关系在预测情境变化和政策更迭的后果上也颇为有用，它告诉我们在各种可能发生但与现存状况不同的 [或者称之为“反事实 (counterfactual)”] 的情况下将会发生什么。举例子来说，在一项调查人的生产能力——即劳动经济学家所谓的人力资本——的研究计划中，我们考察了教育水平对工资的因果效应 [Card (1999) 综述了这个领域里的研究]。在这里，教育水平对工资的因果效应是指个体接受更多教育所带来的工资增加量。一些研究表明，平均而言大学学历的因果效应是工资水平高出 40%，这是个相当大的回报。教育对工资的因果效应对预测上大学的成本发生变化或者加强义务入学法 (compulsory attendance laws) 所导致的收入变化十分有用。因为这一关系还可以从经济学模型中推导出来，所以它也是理论研究的兴趣所在。

作为劳动经济学家，我们非常喜欢将工人作为研究对象来考察因果关系，但是研究因果关系不必一定以个体劳动者为研究对象，企业和国家也可以拿来考虑。关于后者的一个例子就是 Acemoglu、Johnson 和 Robinson (2001) 完成的殖民地制度对经济增长影响的研究。这一研究关心的是从殖民统治者那里继承了更多民主制度的国家后来是否享受到较高的经济增长。对此问题的回答对我们理解历史

以及思考当代发展政策的效果大有深意。比如说,今天我们可能对伊拉克和阿富汗新创建的民主制度是否对其经济发展意义重大有所疑惑。而且,如今民主对经济增长的因果效应远非一目了然,某些东亚经济体就在没有享受完全政治自由化的同时获得了强劲增长,而众多的拉美国家虽然实现了民主化,却没有因此获得较高的经济增长。

常见问题中的第二个是考虑用于捕捉研究对象间因果效应的理想条件下的实验。比如在研究教育水平和工资间关系的例子中,我们可以想象这样一个政策——给那些有潜在辍学倾向的学生一笔奖励以鼓励他们完成学业,然后来研究相应的结果。实际上 Angrist 和 Lavy(2008)就完成了这样的一个实验。尽管这项研究着眼于诸如大学入学率等短期效果,但长期效果可以很好地指向工资^①。在政治制度的例子里,我们可以来一个时空穿梭,回到过去,随机地将不同的政府结构赋予给那些独立之前的殖民地(这个实验很可能只会被拍成电影,而不是得到国家自然科学基金的支持)。

理想条件下的实验通常是假设出来的。即便如此,假想的实验仍然值得我们深思,因为它可以帮助我们挑出那些富有前景的研究主题。作为一名研究者,你可以想象没有预算约束、没有人权委员会因社会正当性而对你的研究进行规制的情况:类似于有丰厚基金支持的 Stanley Milgram,这位心理学家在 20 世纪 60 年代使用具有高度争议的实验设计,打破了关于权威服从性(the response to authority)的常规研究,在今天他的这一冒险行为却很可能会让他失去工作。

为了寻求对权威服从性的理解,Milgram(1963)表明他可以说服实验的被试者去对那些无辜的不断抗议的受害者执行痛苦的电击(这些电击是假的,受害人也是演员扮演的)。这种做法既聪明又极富争议性:有些心理学家声称对他人执行电击的被试者在心理上会受到实验的伤害。尽管如此,Milgram 的研究阐明了这样一点,即便有一些实验最好留在筹划阶段,但我们还是可以对很多实验进行思考^②。如果你在具备理想实验条件的世界里都无法设计出一个实验来回答你的问题,那么在只有适度预算以及非实验调查数据的情况下,你能够得出有用结果的几率就相当渺茫了。对理想条件下的实验进行描述也可以帮助你准确地表达因果问题。在理想条件下进行实验的思考方法还可以凸显那些你希望操控的力量和那些你希望保持不变的因素。

对于那些不能被任何实验回答的研究问题,我们称之为根本无法识别的问题(Fundamentally Unidentified Questions, FUQ)。那什么是根本无法识别的问题

① 在 Angrist 和 Lavy(2008)中,对学生进行奖励来鼓励他们学习以通过大学入学考试,这一实验本身并不增加被实验者的人力资本,但是他们的研究发现被随机支付奖励的学生在 5 年后接受更高等教育的意愿更大,这会增加被实验者的人力资本,从而使他们在长期中获得较高的工资。——译者注

② 后来在一部电视专题节目中 William Shatner 扮演了 Milgram。至今,经济学家尚未获得过这种荣耀,但 Angrist(本书作者之一)对此仍抱有希望。

呢？第一印象告诉我们那些考虑种族和性别不同导致的因果效应问题应该比较接近于根本无法识别的问题，因为我们很难改变被试者的种族和性别，从而很难将这些因素与其他因素隔离分析（将其他因素隔离分析的最好办法是考虑同一个人不同种族或性别下的情况，但这显然不现实，“想象一下你的染色体在出生的时候发生了改变”）。但是在另一方面，经济学家在种族、性别以及劳动市场歧视领域方面的研究集中于是否别人因为你是白人/黑人，男人/女人而对你区别对待（因此被试者究竟是白人/黑人，男人/女人并不重要，重要的是别人对此的看法）。而且，将男人看作女人，将女人看作男人的反事实世界已有很长的历史（比如罗莎琳德化装成古希腊神话中的美少年愚弄莎翁名剧《皆大欢喜》中的每一个人），因此我们对性别的因果效应进行研究并不需要亚当斯式的奇思异想。改变种族的想法显得不那么自然：在《人性污点》（The Human Stain）中，飞利浦·罗斯（Philip Roth，该剧作者）想象出一个以 Coleman Silk 为主角的世界，在那里 Coleman Silk 是一个在职业生涯中以白人身份示人的黑人文学教授。劳动经济学家也总是在设想此类场景。有时候我们甚至要构造这样的一些场景——比如在审计研究（audit studies）中伪造工作申请和简历^①——以推动科学的进展。

将想象推进到研究设计需要走很长一段路，而且想象无法解决所有的问题。譬如说我们对晚一点入学的孩子是否在学校里表现更好这个问题感兴趣。因为七岁孩子的大脑可能比六岁孩子的大脑更适合学习的需要。这个问题来自于事实的政策含义在于有些学区为了提高考试成绩而推迟入学年龄（Deming and Dynarski, 2008）。为了考察延迟入学对学习的影响，典型的做法就是我们随机挑选一些在六岁开始上学的孩子和在七岁开始上学的孩子。我们关心的是：是否如他们小学考试成绩显示的那样，晚一点入学的孩子可以学得更多。具体而言，让我们看一下一年级的考试成绩。

这个问题于是转化为入学年龄对一年级考试成绩的影响，但是问题的关键在于七岁入学的学生年龄本来就比较大。即使纯粹是发育的影响，年长的孩子也倾向于表现出好的成绩。为了避免这个因素对我们考虑问题的干扰，看上去我们应该控制年龄而不是控制学生所在的年级。于是假设我们考虑六岁上学的孩子到二年级时的成绩和七岁孩子在一年级时的成绩，在这个样本里所有孩子的成绩都是在其年龄为七岁时获得的。但是六岁上学的孩子在七岁时已经在学校有一年的时间，呆在学校里的任何有价值因素都可能引起学生成绩的提高（将这一现象称为在学影响）。因此，由于孩子们一直待在学校，所以无法将上面提到的发育影响和在学影响同时剔除以考察纯粹的入学年龄对成绩的影响。问题的核心乃是：对于学生而言，入学时间等于现有年龄减去在校时间（入学年龄 = 现有年龄 - 在校时间，控制发育影响和控制在校时间在学影响相当于使现有年龄和在校时间都不变，那么这个恒

① 最近的一项研究是 Bertrand 和 Mullainathan (2004)，他们考察了面对姓名首字母听上去像白人和听上去像黑人的简历，雇主们的反应有何区别。

等式告诉我们入学年龄也将不变，于是将无法观察不同年龄入学对学习成绩的影响)。上面这个恒等式在由成年人组成的样本中就不再成立了，因此我们可以考察纯粹的入学年龄对成年人的影响，这些影响包括收入水平或者高等教育完成水平(Black, Devereux and Salvanes, 2008)。由于即使在随机实验中也无法获得入学年龄对小学生成绩的影响，因此这个问题就是根本无法识别的问题，即 FUQs。

第三个和第四个研究的常见问题(FAQ)主要涉及完成一项具体研究的细节。第三个问题是这样的：你的识别策略是怎样的？Angrist 和 Krueger(1999)使用识别策略这一术语来描述研究人员运用观察数据(也就是说，不是通过随机实验产生的数据)逼近真实实验的方式。我们再一次回到教育那个例子，Angrist 和 Krueger(1991)把美国学校的义务入学法和学生出生季节所成的交互项作为一项自然实验来估计高中学历对工资的影响(义务入学法要求学生在 16 岁生日以后才能辍学，但学生的出生季节影响学生的辍学率而与学生的能力无关)。本书从第 3 章到第 6 章的内容基本都是在考虑识别策略的概念性框架。

尽管对可靠的识别策略的关注可以作为现代经验研究的象征，但在计量经济学中将真实实验和自然实验相提并论还是经历了一个漫长的历史过程。下面是计量经济学先驱特里夫·哈尔维莫(Trygve Haavelmo, 1944:14)呼吁对两种实验设计进行更为细致的讨论时所讲的一段话：

对于任何进行定量研究的理论而言，实验设计(物理学家称之为“决定性实验”)都是一项基本的附件。在我们构建理论时脑海中通常有一些这样的实验，虽然很不幸的是大多数经济学家无法精细地描述他们的实验设计。如果他们做到了这一点，就会发现其实脑海中的实验可以归为两类，也即(1)将所考虑问题从其他影响中分离出的实验，我们以此来考察特定的真实经济现象是否可以用来验证特定的假设；(2)自然从其巨大的实验室里不断川流而出的实验，对于此，我们主要是作为一个被动的观察者。无论哪种情况，理论的目的都是相同的：主，真实生活带来的幸福。

FAQ 的第四个问题借用了 Rubin(1991)的说法：你进行统计推断的模式是什么？为了回答这个问题，我们需要描述被研究的总体、所使用的样本以及构建标准误时所作的假设。有时候，推断是非常直观的，比如当你使用人口普查微观数据样本来研究美国人口的时候。不过，通常而言，推断则更为复杂，尤其是数据存在聚类问题或者被分组时。本书末章就涵盖了当你需要回答常见问题四时将会出现的实际问题。尽管推断问题往往有相当的技术性而且很难让人兴奋，但是一项经过精良构思而且从概念上来说激动人心的项目，其最终的成功也要依赖于这些统计推断的细节。这一有时令人大感沮丧的事实激发了下面的一段计量经济学俳句，这是一位来自日本的计量经济学博士生 Keisuke Hirano 在完成自己的博士论文时妙手偶得的：

t 统计言笑晏晏，如此迷人

试着将标准差集中——