

计量经济学成长之路

研究设计学习

张毅

2020 年 8 月 12 日

第一章 我的思考

1.1 小喇叭的长期影响

一个可能的方面，文革期间广播节目被暂停，1978 年 11 月 6 日正式开播，节目定位是面向学龄前儿童，因此可以比较在 1978 年 11 月前已上学学生和未上学学生的差异，存在的问题是没有相关电台数据，因此是否可以用安慰剂检验进行研究，比如选一个年龄更高的群体，比较着两类群体在这前后差异不大，因此可以归为是节目的影响。

第二章 处理效应

2.1 处理效应与选择难题

我所理解的处理效应也就是我们通常所说的因果效应，首先需要提到的自然是内生性问题。以就业培训为例，通常的办法是直接比较参加就业培训的人群与未参加就业培训人群的未来收入，发现参加就业人的未来收益低于未参加人群，但并不能说这是由于参加就业培训带来的，这里存在的一个典型的问题就是“自选择”(self selection)。可能的原因是收入高的人群并不需要参加培训。

Rubin and Donald [1974] 提出了“反事实框架”，称为 Rubin Causal Model。虚拟变量 $D_i = \{0, 1\}$ 表示个体 i 是否参与此项目。 D_i 被称为“处理变量”。记个体未来的收入为 y_i 。想知道 D_i 是否对 y_i 有因果作用。

$$f(x) = \begin{cases} y_i = y_{1i} & \text{if } D_i = 1 \\ y_i = y_{0i} & \text{if } D_i = 0 \end{cases} \quad (2.1)$$

y_{0i} 表示个体 i 未参加项目的未来收入，而 y_{1i} 表示个体 i 参加项目的未来收入。 $y_{1i} - y_{0i}$ 即参加培训的因果效应。但是通常只能观察到 $y_{1i}y_{0i}$ 中的一种。

$$y_i = (1 - D_i)y_{0i} + D_i y_{1i} = y_{0i} + (y_{1i} - y_{0i})D_i \quad (2.2)$$

其中 $y_{1i} - y_{0i}$ 为个体 i 参加项目的因果效应。

对内生性问题的一个处理办法是进行随机分组，也就是下一章所讲的随机干预实验。当然，目前大部分学者都是用的是公开数据，那随机分组不满足的情况下，各个准实验方法的目的就在于此。

第三章 随机干预实验：RCT

第四章 工具变量

4.1 父母失业对孩子辍学的影响

The Effect of Parental Job Loss on Child School Dropout: Evidence from the Occupied Palestinian Territories

摘要：Journal of Development Economics 发表的论文利用巴勒斯坦被占地区样本研究父母工作失业对于孩子辍学的影响 [Di Maio and Nisticò, 2019]。

负面的经济冲击如何影响家庭的教育决策？现有研究对发展中国家的关注不够，那不勒斯帕斯诺普大学的 Michele Di Maio 和 Roberto Nisticò 发表在 Journal of Development Economics 上的论文“The Effect of Parental Job Loss on Child School Dropout: Evidence from the Occupied Palestinian Territories”以巴勒斯坦被占领地区样本，利用巴以冲突导致的巴勒斯坦人失业这一外生冲击克服估计中存在的内生性问题，提供了来自发展中国家的证据。1993 年巴勒斯坦与以色列签订了奥斯陆协议，地区局势实现阶段性和平，巴勒斯坦工人得以进入以色列从事低技能工作，在此期间以色列雇用了超过 25% 的巴勒斯坦劳动力，占巴勒斯坦国民收入的 1/6。而 2000 年后地区局势恶化，以色列限制巴勒斯坦劳动力进入以色列境内，导致巴勒斯坦被占地区人口失业，作者通过这一外生冲击研究父母失业对于学生辍学的影响。

研究数据来源于巴勒斯坦劳动力调查数据 (PLFS)，该数据收集了巴勒斯坦 16 岁以上劳动力工作相关的季度信息，结合数据中 2000-2006 年间该数据中 10-15 岁儿童样本，并利用 NGO 组织 B'Tselem 提供的每季度每一万人中由以色列军队造成的死亡人数衡量冲突的严重程度，将其作为巴勒斯坦人口失业的工具变量。识别策略：作者实证研究了父母失业对于学生辍学的影响。对于辍学的定义：若孩子 t 期上学， $t+1$ 期末未上学则定义为辍学；

同样，若受访者 t 期在以色列境内就业， $t+1$ 期在以色列未被雇佣则定义为失业。并控制孩子、家庭特征、地区固定效应以及时间固定效应，进行 OLS 估计。在此基础上利用冲突强度（死亡率）作为巴勒斯坦被占地区人口失业的工具变量。如果冲突直接影响到辍学率那么工具变量的相关假定将无法满足，比如冲突造成的学校基础设施的破坏以及去学校的困难程度进而造成辍学，如果这一假设成立，那么冲突的影响将会对被占地区所有学生产生影响。作者比较了父母在以色列工作与父母在被占地区工作的孩子辍学数据发现，冲突只影响父母在以色列工作孩子的辍学率，进而验证工具变量的有效性；此外，作者发现被占地区的教育系统坚不可摧，冲突地区的学校入学率反而还较前期有所提升，据此得以论证工具变量的有效性。

潜在威胁还在于：第一，学生可能是由于抵抗以色列并非因为父母丢失工作辍学；第二，家庭可能因冲突而决定搬迁，高技能工人的家庭搬到死亡人数较少的地区。对于威胁一，作者通过引用文献佐证入学率与以色列国防军实施的军事措施强度间无相关性，排除了将辍学率与冲突强度联系起来的反馈机制；对于威胁二，作者引用巴勒斯坦统计局数据证明，冲突时期的家庭迁徙不多，且迁徙群体中大部分是出于婚姻而进行的。

在克服内生性问题，并进行一系列稳健型检验（样本扩大、失业定义的变化以及删掉未与以色列接壤的地区样本）后（样本扩大、失业定义的变化以及删掉未与以色列接壤的地区样本）发现的结果表明，父母失业会造成孩子辍学的可能性几率提高 9 个百分点，异质性方面男生与学业表现较差的学生受父母失业影响较大。

评论：这篇文章的在识别部分写的很清楚，也是我第一次能清楚的明白研究设计的重要性。

4.2 饥荒的长期影响

The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment using China's Great Famine

摘要：NBER Working Paper 发表的工作论文采用分位数加工具变量回归，估计出饥荒对幸存者健康的净效应 [Meng and Qian, 2009]。

1959-1961 年间的饥荒造成了大约 16-30 万中国人死亡，幼儿及母胎期间经历过 1959-1961 年饥荒的人群现在正处于其中年时期，现有研究侧重于饥荒对死亡率的影响，但对幸存者的研究却比较缺乏（文章于 2009 年公布最初版本，最近几年研究长期影响的文章多了起来）。世界粮食组织 2002 年估计非洲约有 3800 万人口正面临饥荒的威胁，因此研究饥荒的对幸存者的长期影响具有现实意义。澳大利亚国立大学 Meng Xin 与耶鲁大学 Nancy Qian 撰写的 NBER 工作论文“The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment using China's Great Famine”克服了已有研究存在的识别难题后，估计出幼儿及母胎时期经历饥荒对于个体成年后健康和劳动力产出影响的净效应。

研究饥荒对幸存者的长期影响存在以下识别问题，第一，对于饥荒严重程度的测量有偏差；第二，饥荒的影响可能并不随机，饥荒往往发生在粮食产量不高的地区，背后的影响也许是出生地区繁荣与否对于结果变量的影响；第三，饥荒对于幸存者有选择效应，活下来的人群大部分都是比较强壮的群体，而这也与未来的健康产出高度相关。以上三点都会对估计结果有影响。

针对前两个问题，作者首先采用县级层面的人口队列规模来衡量饥荒严重程度，在此基础上使用 1997 年自然环境和播种面积及不同出生群体上的差异作为饥荒严重程度的工具变量（饥荒往往发生在粮食产量不高的地区）；针对问题三，高分位数的估计结果较为准确的反映了饥荒的真实影响，因此采用分位数回归进行分析。此外，由于人口迁移或者其他外生冲击同时发生而造成估计结果的不准确。当时我国严格禁止跨地区移民，因此不存在由于人口迁移而影响估计结果的问题；其次，这次饥荒是在一个政治比较稳定的背景下发生的，并无其他类似于战争的影响，因此可以识别出饥荒的净影响。

文章的分析结果表明，母胎或者幼儿时期经历饥荒对成年后的身高、体重、BMI、受教育程度以及劳动力供给有很大的负面影响。

第五章 固定效应

第六章 双重差分

第七章 断点回归

7.1 理论

有时处理变量 D_i 完全由连续变量是否超过某个断点所决定。比如，考上大学对工资收入的影响，假设大学分数线为 500 分。

$$D_i = \begin{cases} 1 & \text{if } x_i \geq 500 \\ 0 & \text{if } x_i < 500 \end{cases} \quad (7.1)$$

对于考分在 500 左右的学生 (489、499、500、501)，可以认为他们在各方面都没有系统性差异，尤其能力方面。因此可以估计出 $x = 500$ 附近的因果效应，称为局部平均处理效应 (Local Average Treatment Effect, LATE)

$$\begin{aligned} LATE &\equiv E(y_{1i} - y_{0i} | x = 500) \\ &= E(y_{1i} | x = 500) - E(y_{0i} | x = 500) \\ &= \lim_{x \rightarrow 500^+} E(y_{1i} | x) - \lim_{x \rightarrow 500^-} E(y_{0i} | x) \end{aligned} \quad (7.2)$$

表明

=====

=====

=====

=====

=====

第八章 合成控制

第九章 参考文献

引用用文献的案例在此处 Di Maio and Nisticò [2019]
[Di Maio and Nisticò, 2019] 注意两类的不同

参考文献

- Michele Di Maio and Roberto Nisticò. The effect of parental job loss on child school dropout: Evidence from the Occupied Palestinian Territories. *Journal of Development Economics*, 141:102375, November 2019. ISSN 03043878. doi: 10.1016/j.jdeveco.2019.102375. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304387818309702>.
- Xin Meng and Nancy Qian. The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment using China's Great Famine. Working Paper 14917, National Bureau of Economic Research, April 2009. URL <http://www.nber.org/papers/w14917>. Series: Working Paper Series.
- Rubin and B. Donald. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 1974.