

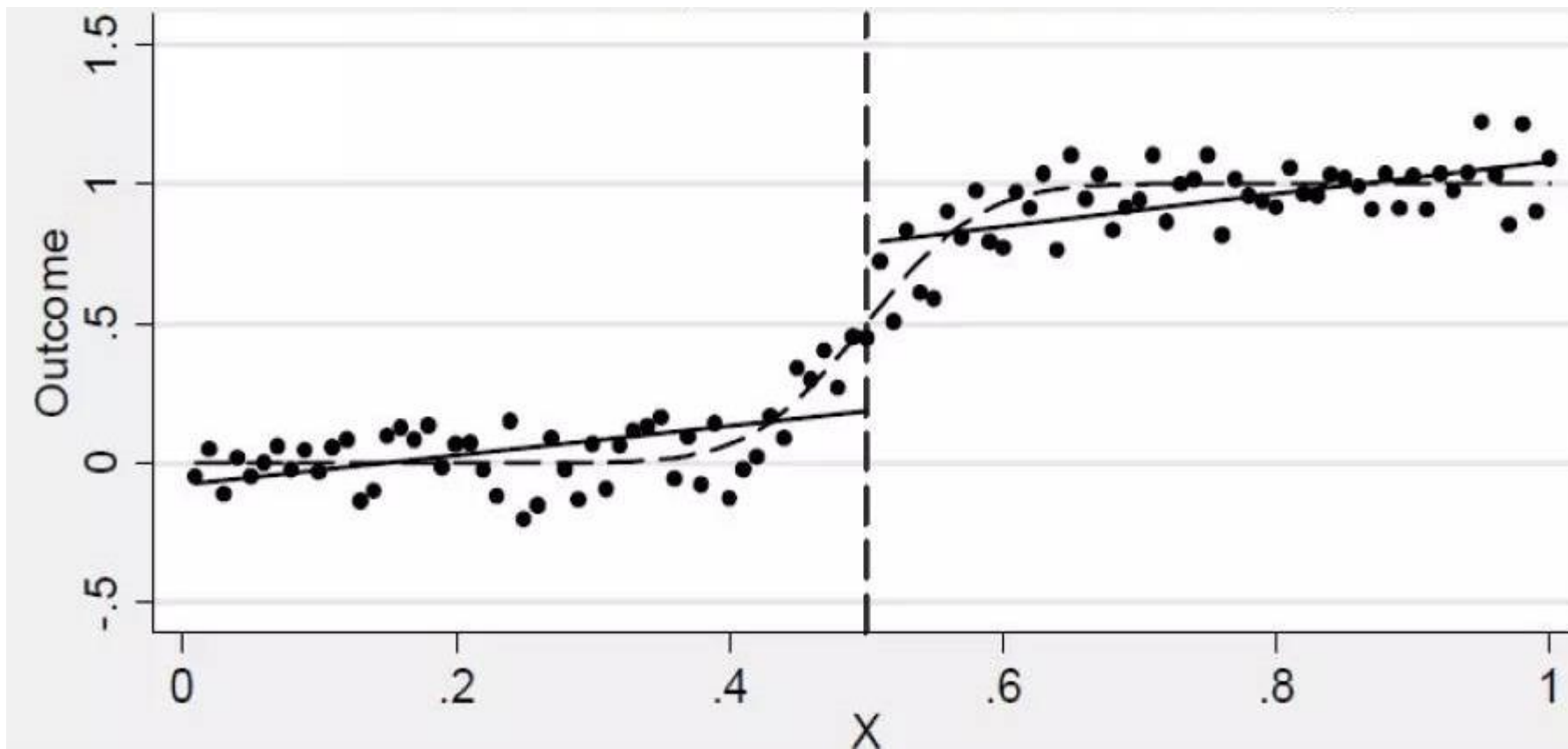
第10章 政策效应评估方法简介

华北电力大学法政系

公管教研室马冉

ncepumaran@163.com

一、什么是政策效应评估？



- ▶ 从科学的意义上来说，评估一项政策的实施效应其背后的基本逻辑应该是：在**给定其他条件相同**的情况下，实施政策后的表现与假定没有实施政策后的表现的差异。

e.g.接受高等教育对个人收入的影响

a.小明—上大学—工资

b.小明—不上大学—工资 反事实参照组

$a-b$ =高等教育对收入的影响

- ▶ 核心问题：用未受到政策干预的人群作为政策干预人群的反事实组，也就是说如果那些受到干预的人群没有受到干预，其结果应该与未受到干预人群的结果是一致的。

► 横向对比案例：评估医疗卫生对国民健康的改善程度

采取随机抽样问卷的方式，询问一个人在过去一周是否去过医院，这个问题会呈现两种状态，如果去过医院，则是受到了医疗的干预，如果没有去过，则是未干预组；同时，让接受问卷的人回答其自评健康程度，健康程度分为5档，1至5分别表示健康程度由差到好。最后，将调查的结果分组汇总起来，就得到如下的统计表格。

表1 横向对比的偏误：医疗

组别	样本数量	平均值	标准差
去过医院	7774	3.21	0.014
没去过医院	90049	3.93	0.003

来源：Angrist and Pischke(2009), P13.

- ▶ 那么，纵向对比可行吗？即简单对比政策前后的差异，并将这种时间上的差异等价于政策效应。

考虑两个因素：

- 时间趋势
- 共同冲击

► 纵向对比案例：2011年实施的房产税试点政策

2011年2月份，为了遏制房地产价格过快上升的势头，我国开始在部分城市进行房产税的试点，选择了上海和重庆两个试点城市。

与其他税种不同，房产税是在持有环节征收，等价于增加了房屋的持有成本，理论上与提高利率的作用渠道是类似的，可以挤出一部分投资性的需求，进而发挥调节房价的作用。

结果，决策层和舆论界都认为房产税试点没有起到遏制房价的作用。重庆的房产税试点从2011年2月开始，但是重庆房价并没有在试点时出现过下降，相反，试点当年的房价反而上涨了11%，看上去重庆的房产税政策并没有使得重庆的房价下降，因此，就得出房产税无效的结论。

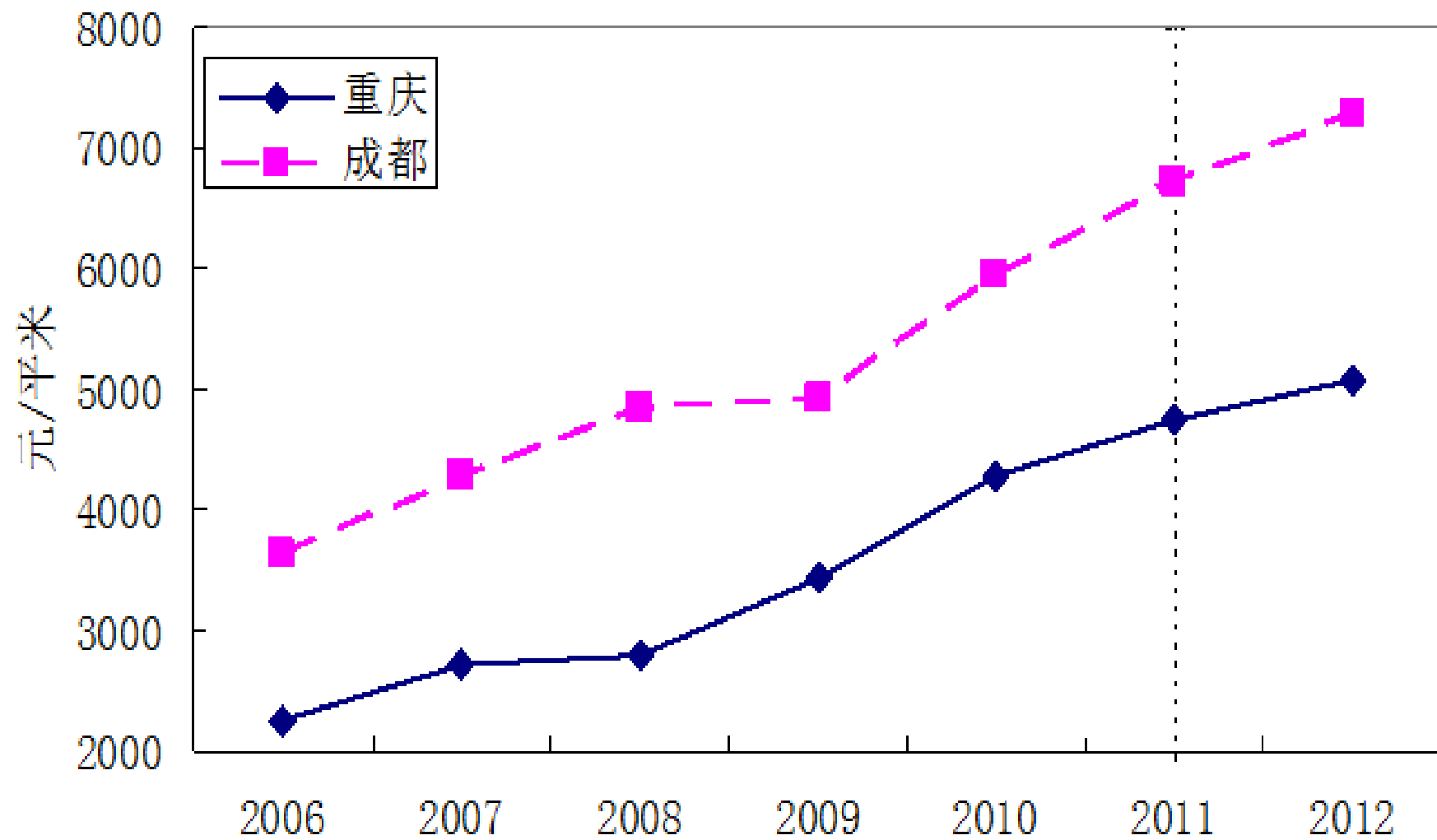


图 1 纵向对比的偏差：房产税

- ▶ 政策评估的科学方法，其质是要找到反事实参照组。
- ▶ 反事实参照组要满足两个基本条件：
 - 一是要保证呈现反事实的特征，即一旦实验组没有受到政策影响，实验组的结果应该与反事实组是完全一致的；
 - 二是反事实参照组不能受到实验组的影响，即那些受到政策干预的人群不能把这种影响传递给其他人。

理论上是允许实验组内部的互相影响，但如果这种影响外溢到参照组，无论这种外溢是主观还是客观原因，都会导致对政策效应估计的偏误，一般情况下都会低估政策的效应。

► 如何构造反事实参照组？

一、数据筛选

经典案例：同卵双胞胎

适合用于就业、收入等方面的应用，但同卵双胞胎通常位于同一辖区，在同一政策覆盖下，因此应用性较差。

二、随机干预试验

医疗卫生案例：针对那些有就医需求的所有人群，通过抛硬币的方法来决定是否可以去医院，例如当抛硬币为正面时，就可以上医院，如果为反面，则不能上医院，由于抛硬币的正反面结果是一个随机冲击，因此对那些有就医需求的人群通过抛硬币的方法，就是一种随机分组，我们事后来评估有就医需求并去医院（正面硬币）和有就医需求并没有去医院（反面硬币）的健康差异，就是医疗对健康的实际改善程度。

伦理问题；随机性难以满足

二、政策的分类

► 根据政策实施的过程，可以将一个国家或部门的政策分为三类：

► 第一种是“先行先试”的政策

决策层无法准确预判其社会和经济影响，政策的容错空间小，需要挑选一些地区或者行业做政策试点，如果试点的效果比较理想，则可以在全国层面推广，反之则不推广。

e.g. 营改增改革

2012年1月1日在上海挑选了“1+6”个服务业行业首先进行试点，然后再逐步将这“1+6”行业推广到八省市和全国，再逐步扩围至电信、邮政等行业，2016年5月开始覆盖房地产业、建筑业、金融业、生活服务业最后四个行业，完成了从先行先试到全范围推开的过程。

➤ 第二种是“一刀切”政策

“一刀切”的政策有两个要点：一是其门槛是非常清晰的，超过（或者低于）该门槛才会进入到政策范围；二是符合政策条件的那些群体会得到较多的资源支持。

e.g.1994年实施的“八七扶贫攻坚计划”

1992年的人均GDP低于400元的县才能进入贫困县，反之，如果一个县的人均GDP超过400元，即使仅为401元，也会失去了贫困县的资格。针对那些贫困县，中央财政在对地方进行转移支付时，会给予额外的倾斜。

➤ 第三种是一次性推开的政策

带有自上而下的特征，决策层对该政策的效应有充分的论证和把握，能够合理预期政策带来的冲击，并且做好了相应的备案，因此可以选择某一时点在全国范围内一次性推开。

e.g.1994年实施的分税制改革

20世纪90年代初，财政包干制下的中央财政入不敷出，财政部只得“打白条”，向每个省借款从1000万到1亿不等。自1994年开始，全国所有省级地区与中央财政采取以税种划分的财政制度，分税制针对全国所有地区实施相同的分成规则，时间节点清晰明朗。

三、政策效应评估的四种主流方法

► 1. 双重差分法 (Difference-in-difference, DID)

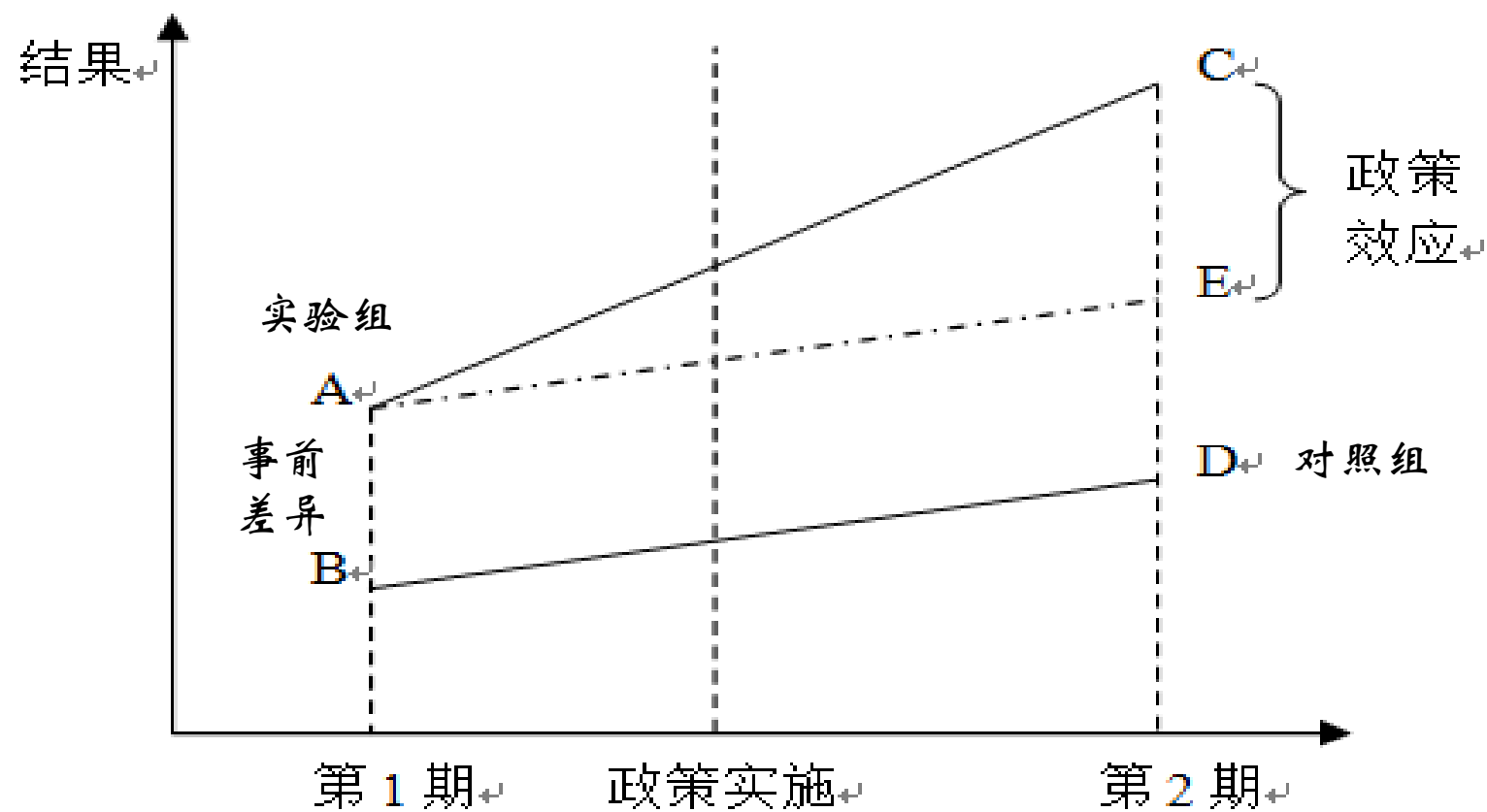


图2 双重差分法的图示

► 双重差分法的优点

- 计量模型简单易用，回归估计方法成熟
- 避免政策作为解释变量存在的内生性问题
- 对政策效应的估计科学、准确
- 高级

► 双重差分法的局限性

- 数据要求高
- 个体时点效应未得到控制
- 未考虑个体所处的环境对个体的不同影响

► 双重差分的稳健性检验

➤ 平行趋势假定

➤ 处理变量对产出变量作用机制的排他性

A.安慰剂检验，即通过虚构处理组进行回归，具体可以：

a) 选取政策实施之前的年份进行回归；

b) 选取已知的并不受政策实施影响的群组作为处理组进行回归；

B.利用不同的对照组进行回归，看研究结论是否依然一致；

C.选取一个完全不受政策干预影响的因素作为被解释变量进行回归。

► 例：“跑部钱进” “跑部进金”（范子英，2014）

1998—2007年间国务院所有部委部长的个人信息数据，将部长的出生地与相应地级市进行匹配，以2003年政府换届为政策实施点，在2002年年底至2003年两会期间，有17个部委更换了部长，将在2002年之前没有部长、之后有部长的地级市作为实验组，将一直没有部长的地级市作为控制组。

结果显示在部长换届之前，两组城市获得的专项转移支付没有差异，在换届之后，有部长的地级市获得了更多的转移支付。平均来说，部长会使得出生地城市的转移支付增加28%，如果是重要部委，增幅高达130%。从图3也可以看出，这种效应呈现先增后减的“U”型特征。

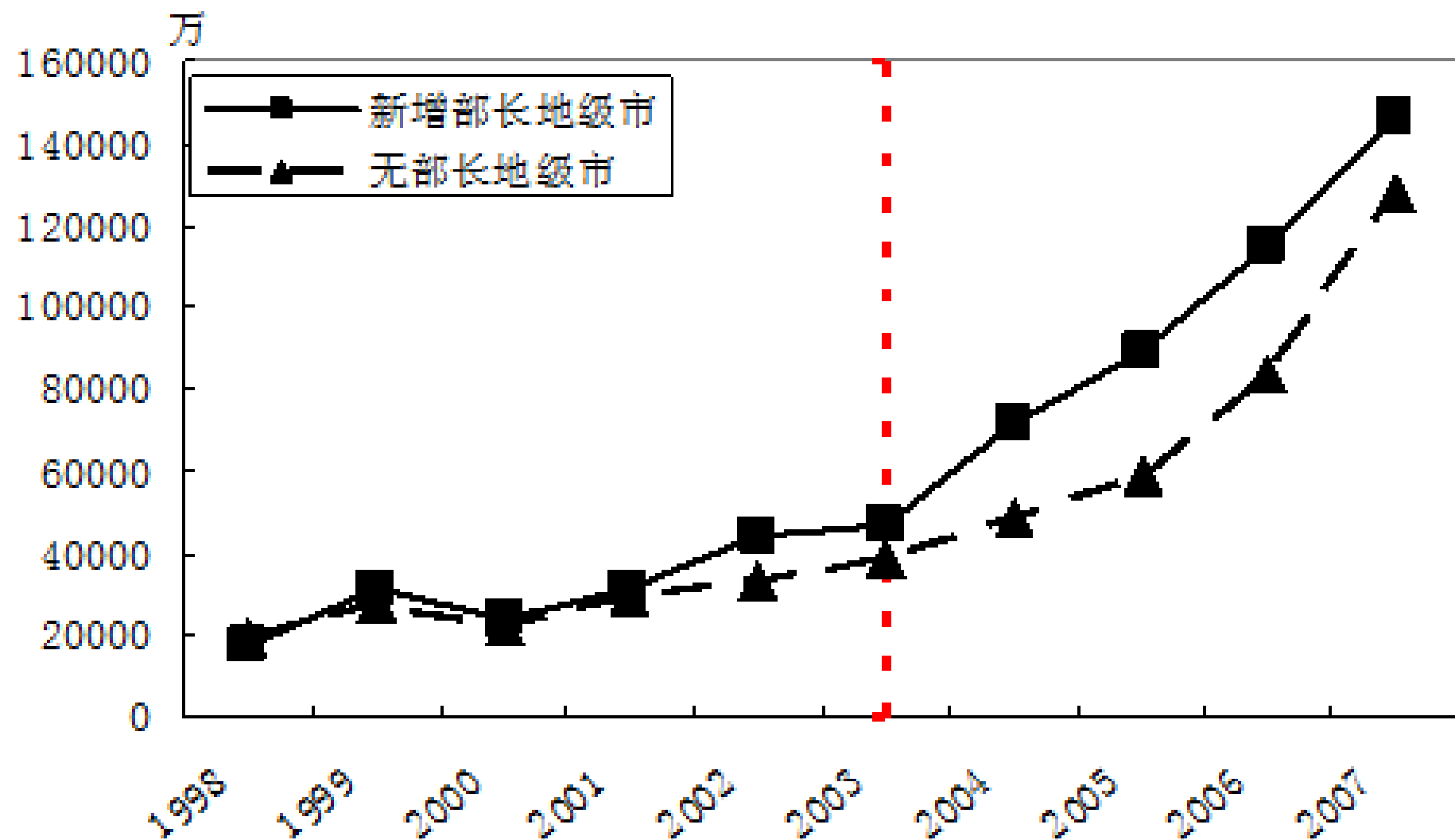


图3 倍差法案例：跑部钱进

► 2.匹配 (Matching)

匹配方法假定，不可观测因素不影响个体是否接受政策干预的决策。因此，对每一个实验组个体而言，可以根据可观测特征为其选择一个控制组个体构成反事实。

根据选择控制组时匹配方法的不同，Matching又可分为协变量匹配(covariantMatching, CVM)和倾向得分匹配(propensityscorematching, PSM)等。其中，CVM涉及多个协变量，会导致“维度灾难”、计算过于复杂等问题。PSM可以克服CVM的劣势，成功降维，应用更为广泛。

► 倾向得分匹配 (PSM)

- 适用：使用非实验数据或者观测数据进行干预效应评估
- PSM解决的主要问题是样本偏差问题。E.g.评估药效
- 倾向得分的定义是一个个体进入处理组的概率，为什么要引入这个概念？

因为如果针对多个可观察特征 X 进行对比匹配非常困难，所以，可将多维协变量 X 用一个一维变量——倾向得分 $P(X)$ 来代替，这样，研究者就只需要对单一的倾向得分变量进行匹配，从而大大减少了匹配的困难。

这个方法最早由Rosenbaum and Rubin(1983)提出。

► 倾向得分匹配的优点：

- 不需要对可观测因素的条件均值函数和不可观测因素的概率分布进行假设

► 倾向得分匹配的局限：·

- 极强的前提假设——强可忽略性假设
- 样本流失
- 数据量要求极大
- 无法处理由不可观测变量带来的内生性问题
- 结果的稳健性受到多种挑战

► 例：父母外出务工对农村留守儿童学习成绩的影响（陶然,周敏慧.2012）

在父母外出对留守儿童学习成绩影响的研究中，内生性问题主要来源于**父母是否外出打工的选择并不是随机的**。该决策会受到父母自身状况、家庭条件，以及孩子学习、身体状况等的影响，甚至父母外出打工决策本身就受到孩子学习成绩左右（吕开宇等，2008；东梅，2011）。由于内生性问题，简单地将留守儿童与非留守儿童成绩对比而得到回归系数不具有统计上的一致性，因而其可信度较低。

具体步骤如下：首先，根据那些可观测的个人和家庭特征，估计出每个孩子成为留守儿童的概率，得到其倾向得分（Propensity Score）；然后寻找与留守儿童的倾向得分最相近的非留守儿童（Nearest Neighbor PSM）作为其反事实，之后比较两组间学习成绩的差异，再对计算出来的差异取均值，得到父母外出对留守儿童成绩的平均影响效应。

► 3. 断点回归法 (Regression Discontinuity, RD)

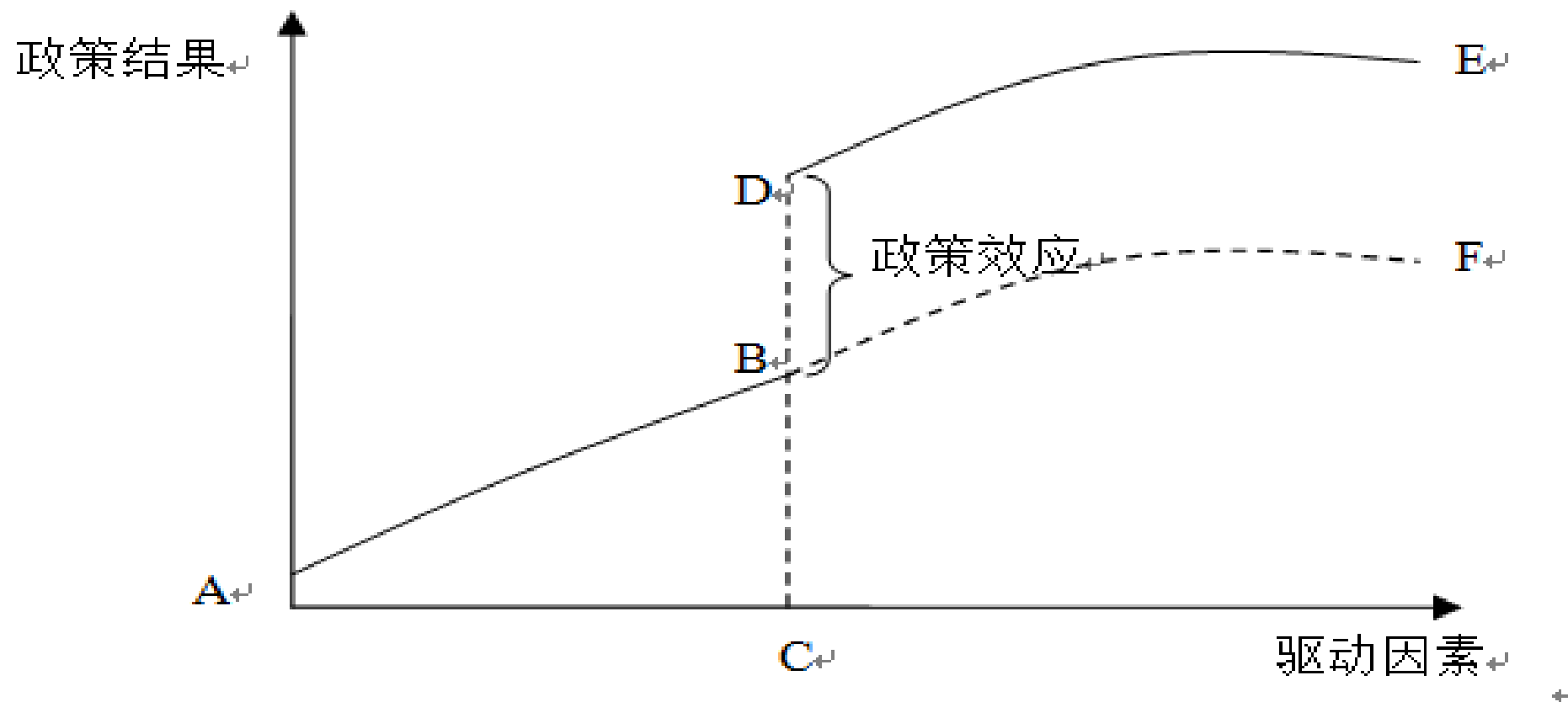


图 4 断点评估法的图示

▶ 三个基本要求：

- “一刀切”政策执行彻底
- 选择门槛的依据要素必须是个体不能改变的
- 影响政策结果的其他因素不能在门槛附近出现跳跃

▶ RDV.S.DID

- RD：选取其他特征相似的组，考察临界值区间上下不同
- DID：比较两组存在差异的群体，但是该差异的影响随着时间变化是恒定的

► 断点回归的优势

- 仅次于随机试验的，能够有效利用现实约束条件分析变量之间因果关系的实证方法。

► 断点回归的局限性

- 在使用断点回归时，如果其他协变量也存在着“中断”的情况，则不清楚是由于其他变量还是我们所关心的强制变量所导致的。
- 非混淆假设条件严格。断点回归方法假设研究对象是同质的或近似同质的，即被放置对照的个体若放在实验组与放置在实验组的个体产生的效应是一样的，但在实际中很难保证，如若产生异质性反应，则估计结果是有偏的。
- 断点回归衡量的是在临界值附近的局部平均效应，不是一个整体的平均效应，很难推广到整体研究中。

► 例：污染的健康损失效应（陈玉宇，2013）

最直接的方法就是选某一年城市层面的数据，例如PM2.5和预期寿命，看看这两个指标是否负相关。在现实中这个系数都是正的，最典型的原因就是忽视其他因素的作用，例如大城市往往也是医疗条件更好的地方，医疗条件会改善健康。即使我们考虑到城市层面的一些特殊因素，简单的统计和经济学方法依然不能准确估计出污染对健康的危害。

以淮河为界的政策是“一刀切”的，在未实施集中供暖政策之前，淮河北南的城市之间差异很小，特别是淮河沿岸的城市之间几乎没有差异，因此淮河南岸的城市可以作为北岸城市的反事实参照组，这句话也就意味着，假设没有集中供暖这一政策，淮河北岸城市的污染水平和人均预期寿命，和南岸城市是完全相同的。当我们发现政策实施之后，两者之间的污染和预期寿命均有差异，那预期寿命的差异就只可能是由污染导致的。

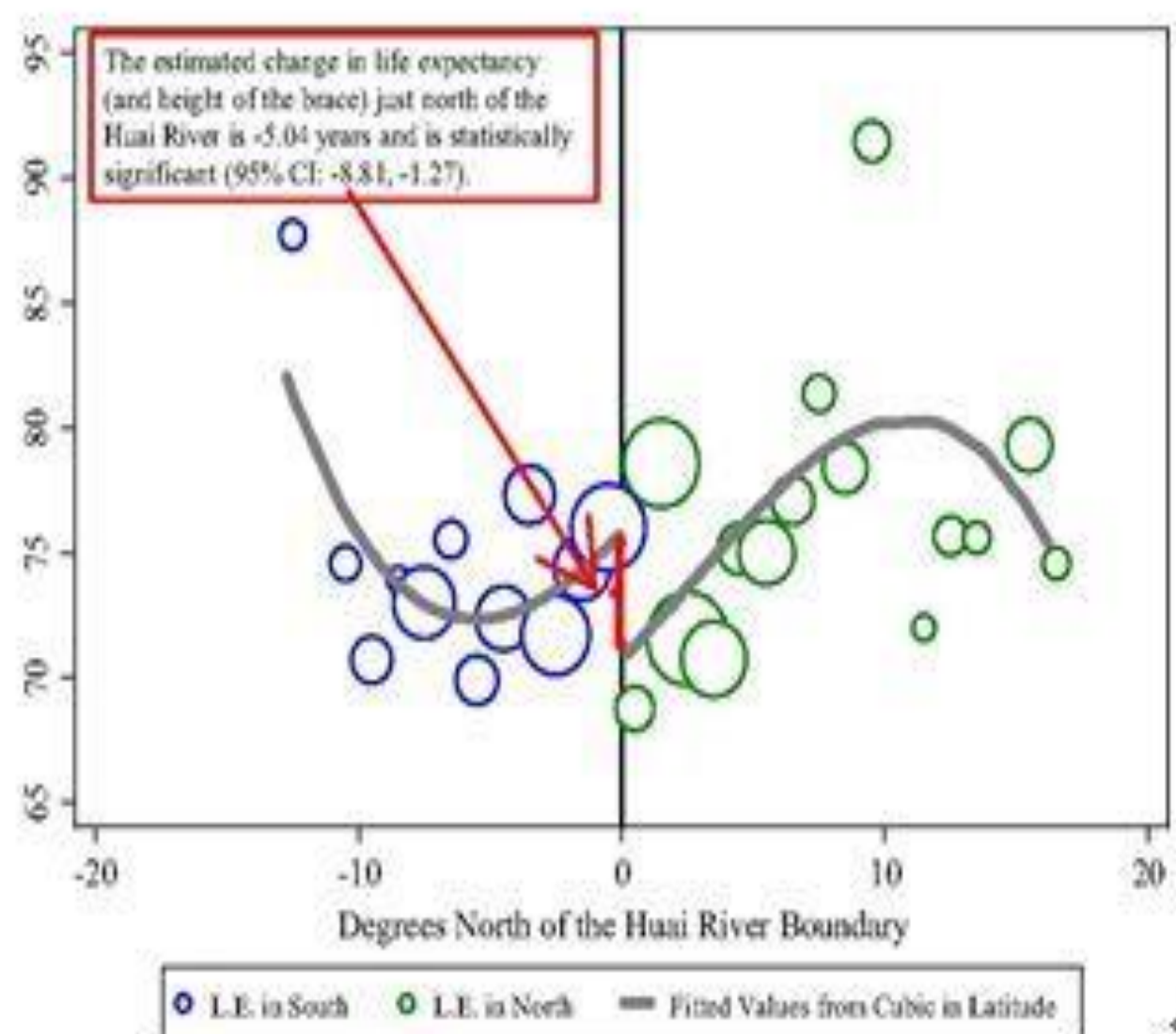
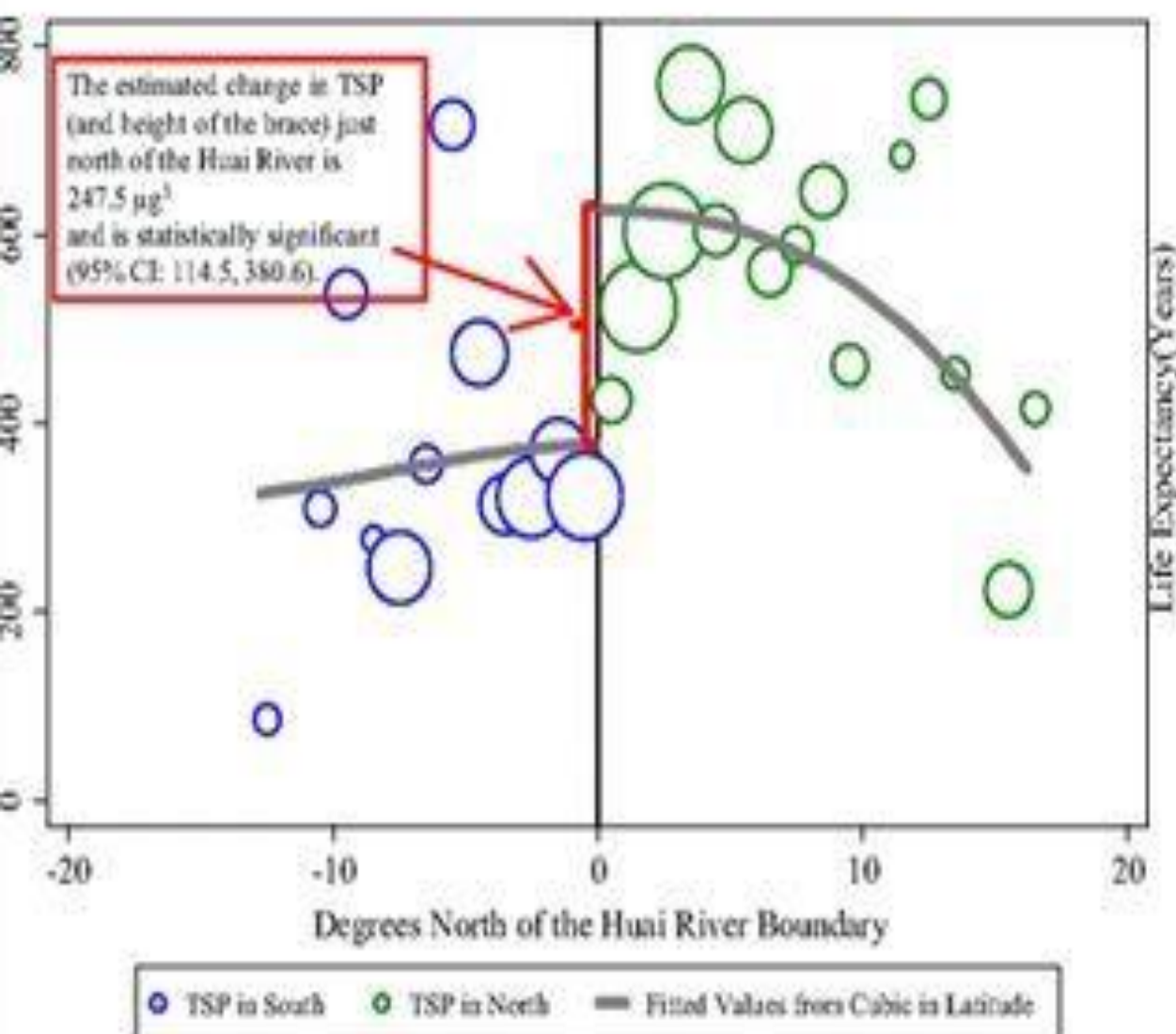


图5 断点评估法案例：污染和健康

资料来源：Chen et al. 2013, Fig. 2 and Fig. 3。

► 4.工具变量法 (IV)

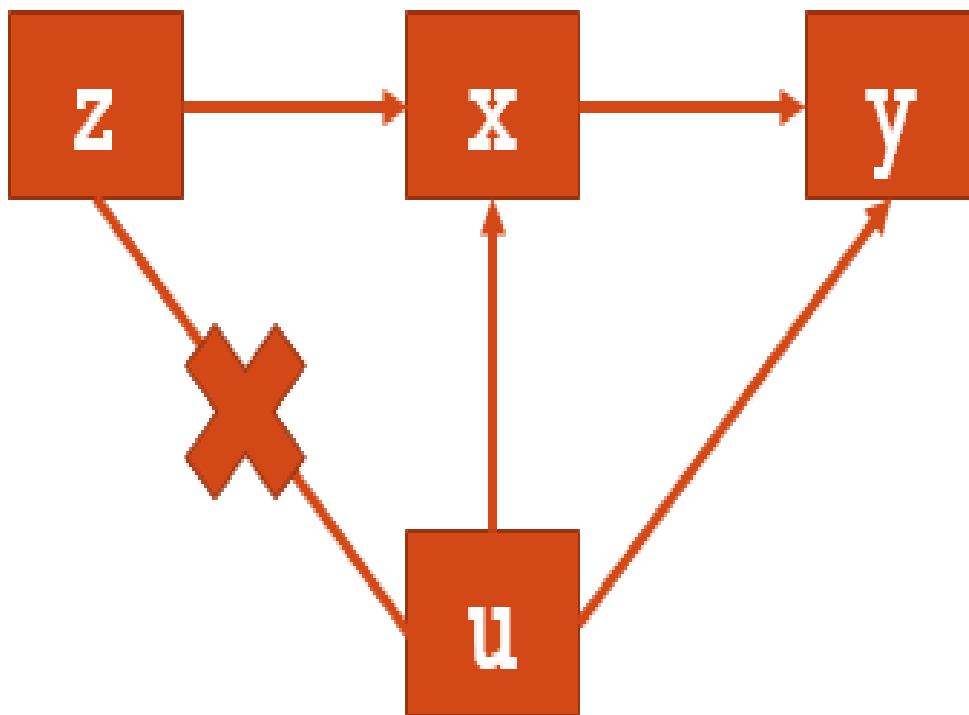
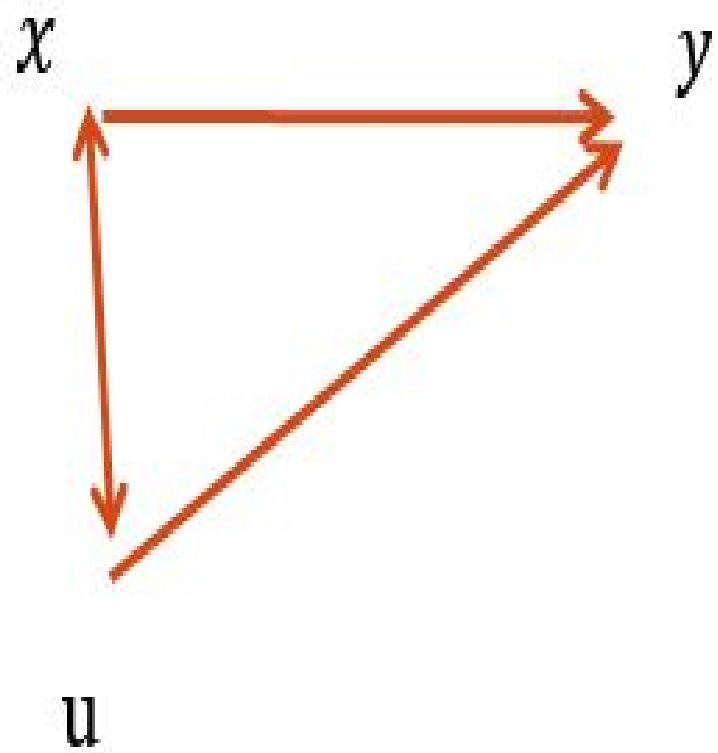
工具变量本质上是来解决内生性问题的

比如：研究一个人的工资水平wage由什么决定，我们提出由：性别male, 年龄age, 智商IQ三个因素决定，因此，我们有：

$$wage = \beta_0 + \beta_1 age + \beta_2 male + \beta_3 IQ + u$$

这里有一个问题，即：IQ由于存在测量误差，因此和u相关，这会导致IQ的估计系数有偏。

这时需要找一个与模型中随机解释变量高度相关，但却不与随机误差项相关的变量，那么就可以用此变量与模型中相应回归系数得到一个一致估计量，这个变量就称为工具变量，这种估计方法就叫工具变量法。



► 作为工具变量，必须满足下述四个条件：

- (1) 与所替的随机解释变量高度相关；
- (2) 与随机误差项不相关；
- (3) 与模型中其他解释变量不相关；
- (4) 同一模型中需要引入多个工具变量时，这些工具变量之间不相关。

► 工具变量法的缺陷：

- (1) 工具变量的选择问题。
- (2) 如果个体对于政策的反应不同，只有当个体对政策反应的异质性并不影响参与决策时，工具变量才能识别处理效应。

► 例：制度软环境与经济发展（董志强，2012）

政府致力于塑造营商软环境是为了经济发展，同时其政策选择和制度设计等行为又受制于经济发展的状态和水平。这意味着营商软环境因而可能部分地内生于经济发展过程。为了检验营商软环境对经济发展的单向影响关系，必须寻找到影响当前营商软环境的外生因素作为工具变量。

选择的工具变量是各样本城市开埠通商的历史。开埠通商历史越长，则开埠时间越早，受西方的影响就越早越深；而且从我国工商业发展的历史来看，开埠通商越早的城市其现代工商业也孕育得更早。这意味着开埠通商历史越长的城市，很可能因为这层历史原因而具有更好的营商软环境。

