

# 遗漏变量偏误估算及 Stata 实现

王非

武汉

2019.06.16

# $X$ 对 $Y$ 的影响

- 从一元线性回归开始
  - $Y = \alpha + \beta X + u$
  - $Y$ : 工资;  $X$ : 教育水平
  - 遗漏了许多重要变量
    - $X$  与  $u$  相关  $\longrightarrow \beta$  的估计有偏误

# 传统处理方法

- 控制变量
  - 常有一些重要变量无法准确度量
  - 加减自变量，检验  $X$  系数的稳健性：局限性
- 更高级的方法
  - 工具变量，.....
- 传统处理方法无能为力，该怎么办？

# 退而求其次

- 如果不处理遗漏变量偏误问题，系数偏误能有多大？会影响基本结论吗？
- Emily Oster (2017), Unobservable Selection and Coefficient Stability: Theory and Evidence, *Journal of Business & Economic Statistics*.

# 基本思路

- 在一元线性回归  $Y = \alpha + \beta X + u$  中，有两类遗漏变量
  - 研究者可以观测并控制起来的自变量： $W$
  - 研究者观测不到的自变量： $Z$
- 两个关键指标
  - $Z$  与  $X$  的关联程度 除以  $W$  与  $X$  的关联程度： $\delta$
  - 在  $Y = \alpha + \beta X + \gamma W + u$  基础上，如果额外控制  $Z$ ， $R$  平方会增加多少？
- Oster (2017) 的建议
  - $\delta = 1$
  - $R$  平方增加 30%

# Stata 演示

命令：`psacalc`

# 结果展示格式

	[1]
一元线性回归教育水平系数	0.743*** (0.046)
$R$ 平方	0.106
控制其他变量后教育水平系数	0.722*** (0.046)
$R$ 平方	0.112
$R$ 平方 * 1.3	0.146
$\delta = 1$ 时教育水平系数	0.599*** (0.065)
教育水平系数为 0 时的 $\delta$ 值	1.980

# 补充提醒

- Oster 的方法只起兜底作用，最理想的还是尽可能估算出准确的系数
- 在数据分析中，如果对系数识别方法没有足够信心，Oster 检验可以作为补充，以检验结果稳健性