

## 2021 中级计量经济学作业 2

---

1. (关于分块回归) 已知  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + \varepsilon$ . 证明 OLS 回归系数估计  $b_K$  满足

$$b_K \rightarrow \beta_K + \frac{\text{Cov}(\tilde{x}_K, \varepsilon)}{\text{Var}(\tilde{x}_K)},$$

其中  $\tilde{x}_K = x_K - L(x_K | 1, x_1, \dots, x_{K-1})$ ,  $L(x_K | 1, x_1, \dots, x_{K-1})$  是  $x_K$  在  $1, x_1, \dots, x_{K-1}$  上的 Linear Projection. (假设所需的秩条件和大数定律满足)

2. 在同方差假设下, 比较 OLS 回归和 2SLS 回归的渐进方差, 说明哪一个更加有效 (more efficient). (提示:  $\mathbf{A} - \mathbf{B}$  为半正定等价于  $\mathbf{B}^{-1} - \mathbf{A}^{-1}$  为半正定.)
3. 考虑如下关于吸烟是否影响新生儿体重的模型:

$$\log(\text{bwght}) = \beta_0 + \beta_1 \text{male} + \beta_2 \text{parity} + \beta_3 \log(\text{faminc}) + \beta_4 \text{packs} + \varepsilon,$$

其中  $\log(\text{bwght})$  是新生儿体重的对数,  $\text{male}$  是新生儿是否为男孩的虚拟变量,  $\text{parity}$  是新生儿出生的顺序,  $\text{faminc}$  是家庭收入,  $\text{packs}$  是孕妇在怀孕期每天吸烟的平均数.

1.  $\text{packs}$  会否和  $\varepsilon$  相关? 为什么?
2. 假设你取得了样本中妇女所在省的香烟平均价格  $\text{cigprice}$  作为  $\text{packs}$  的工具变量. 讨论  $\text{cigprice}$  能否满足工具所需的两个条件.
3. 用 BWGHT.DTA 数据估计模型. 先用 OLS. 再用 2SLS, 用  $\text{cigprice}$  作为  $\text{packs}$  的工具变量. 结果是否有很大的不同? (可直接附上 Stata 回归结果, 无需考虑格式)
4. 作第一阶段回归,  $\text{cigprice}$  和  $\text{packs}$  的相关性是否足够强? 这个结果如何影响了 2SLS 回归?