

计量经济学 第三次作业

2020 年秋

注意事项:

1. 请于 **2020 年 12 月 4 日下午 5 点**前提交电子版至网络学堂，不接受以其他形式提交的作业。对作业提交有疑问请发邮件联系助教彭鹭 (pengl.18@sem.tsinghua.edu.cn)。
2. 上机作业的数据集和代码请从网络学堂 → 课程文件 → 习题集下载。
3. 请将所有作业的答案生成一份 PDF 文档（手写扫描或者打印皆可）提交。在提交前，请确认文件是否可以正常打开。
4. 欢迎在提交作业前参加助教答疑，也可以和同学讨论作业思路，但请提交你自己写的答案。

[沃德 (Wald) 估计量] 考虑简单回归模型

$$y = \beta_0 + \beta x + u. \quad (1)$$

(1) 其中 x 是内生变量. 令 z 为 x 的工具变量, 同时 z 只能取 0 或者 1 两个值. 令 \bar{y}_0 和 \bar{x}_0 代表 $z_i = 0$ 那部分样本中 y_i 和 x_i 的样本均值, \bar{y}_1 和 \bar{x}_1 代表 $z_i = 1$ 那部分样本中 y_i 和 x_i 的样本均值. 证明 IV 估计量 $\hat{\beta}_{IV}$ 可以写成如下形式 (20 分)

$$\hat{\beta}_{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}.$$

(2) 研究者想检验儿女数目对女性就业率的影响。为此研究者收集了一些母亲的样本，并用第一胎是否为双胞胎作为儿女数目的 IV。在你看来，这是一个合理的 IV 吗？换句话说，请解释在满足什么条件的情况下，双胞胎可以做为儿女数目的 IV (7 分) 如果这些条件满足，请解释你将如何利用这个 IV 构造估计量 (8 分)

[异方差性：GPA 的决定因素] 某大学为了研究学生成绩的决定因素收集了在读学生的数据，包含以下变量：

- colGPA 大学 GPA
- hsGPA 高中成绩
- ACT 入学前标准化考试成绩
- skipped 每周翘课的数目
- PC 是否拥有个人电脑

上机作业的数据集和对应的 Stata do file 可以从网络学堂 → 课程文件 → 习题集下载，Stata do file 文件名为 “gpa.do”，数据集文件名为 “gpa.dta”。注意：

- **此题不要求上传 Stata do file**
- 请直接下载代码后研究清楚每个步骤的作用（遇到看不懂的代码可以利用 help 指令），并在自己的计算机上运行结果，然后回答下列问题。

(1) 用 OLS 估计下面的模型并给出回归结果（请报告同方差假设下的标准误，附上回归结果的截图）（5 分）

$$\text{colGPA} = \beta_0 + \beta_1 \text{hsGPA} + \beta_2 \text{ACT} + \beta_3 \text{skipped} + \beta_4 \text{PC} + u.$$

(2) 请画出残差图（纵坐标为残差，横坐标为拟合值，请使用 `rvplot` 指令，并附上图片。）在你看来，直观上，这个数据是否存在异方差性，为什么？（5 分）

(3) 使用 BP 检验 ($\hat{u}^2 \sim \hat{y}$) 验证同方差性。H0: 不存在异方差。在 5% 的显著性水平下，是否可以拒绝原假设？（5 分）

(4) 下面我们用

$$\hat{u}^2 = \alpha_0 + \alpha_1 \hat{y} + \alpha_2 \hat{y}^2 + v$$

估计 WLS 的权重。请估计上述方程并求拟合值 \hat{h}_i 。在你的回归中， \hat{h}_i 是否都是正数？（5 分）

(5) 使用上一问中得到的 $1/\hat{h}$ 作为权重，用 WLS 估计（1）中的模型并给出回归结果（请报告同方差假设下的标准误，并附上回归结果的截图）（5 分）

(6) 对比 OLS 和 WLS 的回归结果，skipped 和 PC 的估计值变化大吗？统计显著性如何？（5 分）

(7) 再一次用 WLS 估计 (1) 中的模型, 请使用相同的权重, 但是这次报告异方差稳健的标准误并附上截图。新的标准误和 (5) 中的标准误相比有很大变化吗? (5 分)

[工具变量: 教育的回报率] 卡德 (Card, 1995) 在一篇论文中¹用住址附近是否有大学作为受教育年限的工具变量并估计教育对收入的影响。我们使用相似的数据复制这篇文章的结果。数据集中包含如下信息:

- `lwage` 工资的自然对数
- `educ` 受教育年限
- `nearc4` 居住地附近有大学 (1 表示附近有大学, 0 表示没有)
- `exper, exper2` 工作经验及其平方
- `smsa` 是否居住在大城市 (1 表示居住在大城市, 0 表示住在农村)
- `motheduc` 母亲的受教育水平

上机作业的数据集和对应的 Stata do file 可以从网络学堂 → 课程文件 → 习题集下载, Stata do file 文件名为 “card.do”, 数据集文件名为 “card.dta”。注意:

- 此题不要求上传 Stata do file
- 请直接下载代码后研究清楚每个步骤的作用 (遇到看不懂的代码可以利用 `help` 指令), 并在自己的计算机上运行结果, 然后回答下列问题.

(1) 用 OLS 估计下面的模型并给出回归结果 (附上回归结果的截图) (5 分)

$$\text{lwage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper2} + \beta_4 \text{smsa} + \beta_5 \text{motheduc} + u.$$

(2) 我们考虑用 `nearc4` 作为 `educ` 的 IV。请问是否满足工具变量的相关性? 为什么? (5 分)



(3) 在你看来, `nearc4` 是否满足工具变量的外生性? 为什么? (5 分)

(4) 用 `nearc4` 作为 `educ` 的 IV, 并用 2SLS 估计上述模型, 附上回归结果的截图 (5 分)

¹Card, David. “Using Geographic Variation in College Proximity to Estimate the Return to Schooling”, 1995

(5) 2SLS 的估计值比 OLS 更大吗? 2SLS 估计值的方差更大吗? 两者的显著性水平有什么区别? (5 分)

(6) 对模型进行内生性检验, 报告你的检验结果。在 10% 的显著性水平下, 是否可以拒绝原假设 (原假设: `educ` 是外生的)? (5 分)