

本人学号_____

批改人学号_____

成绩_____

特注:

- (1) 你可以与同学共同讨论，但最后必须亲自完成作业。如发现抄袭，将以零分计。
- (2) 如果作业中写了姓名，则该作业无效。
- (3) 只接受纸质版（回归结果须打印），不接受电子版。手写时请注意书写端正。
- (4) 为了公平起见，如果逾期未交作业，将不再接受补交作业。
- (5) 请装订作业，或在每页都写上学号，以避免散落遗失。

Problem Set 2

1. 假设随机变量 y 的期望为 μ ，抽样得到其 iid 随机样本 $\{y_1, \dots, y_n\}$ ，记样本均值为

$$\bar{y} \equiv \frac{1}{n} \sum_{i=1}^n y_i。$$

- (1) \bar{y} 是 μ 的无偏估计。 \bar{y}^2 是否是 μ^2 的无偏估计？
- (2) \bar{y} 是 μ 的一致估计。 \bar{y}^2 是否是 μ^2 的一致估计？

2. 考虑有关啤酒月消费量的线性模型：

$$beer = \beta_1 + \beta_2 inc + \beta_3 price + \beta_4 educ + \beta_5 female + \varepsilon$$

其中， $E(\varepsilon | inc, price, educ, female) = 0$ ， $Var(\varepsilon | inc, price, educ, female) = \sigma^2 inc^2$ 。
对此模型进行变换，使得变换后的扰动项为同方差。

3. 房屋的价格如何决定？一种理论认为，房价由房屋性能所决定，称为“特征价格法” (hedonic price)。数据集 `hprice2a.dta` 包含美国波士顿 506 个社区的房屋中位数价格的横截面数据。考虑以下特征价格回归：

$$lprice_i = \beta_1 + \beta_2 lnox_i + \beta_3 ldist_i + \beta_4 rooms_i + \beta_5 stratio_i + \varepsilon_i$$

其中， $lprice$ 为房价的对数， $lnox$ 为空气污染程度的对数， $ldist$ 为社区到就业中心距离的对数， $rooms$ 为房屋的平均房间数， $stratio$ 为社区学校的学生-教师比例，下标 i 表示社区 i 。

- (1) 使用普通标准误进行回归，并评论解释变量系数的符号、统计显著性及经济意义。
- (2) 使用稳健标准误进行回归。稳健标准误与普通标准误差别大吗？
- (3) 使用稳健标准误，以 5% 的显著性水平，检验 “ $H_0: \beta_3 = \beta_5$ ”。
- (4) 使用稳健标准误，以 5% 的显著性水平，分别检验 “ $H_0: \beta_4 = 0.31$ ” 与 “ $H_0: \beta_4 = 0.30$ ”。

4. 使用数据集 `gasoline.dta` 估计美国 1953-2004 年的汽油需求函数。考虑如下回归：

$$lgasq_t = \beta_1 + \beta_2 lincome_t + \beta_3 lgasp_t + \beta_4 lpnc_t + \beta_5 lpuc_t + \varepsilon_t$$

其中，被解释变量 $lgasq$ 为人均汽油消费量的对数，解释变量 $lincome$ 为人均收入对数， $lgasp$

为汽油价格指数的对数, $lpnc$ 为新车价格指数的对数, $lpuc$ 为二手车价格指数的对数。

- (1) 使用 OLS 进行估计此方程。评论各变量系数的符号、显著性与经济意义。
- (2) 计算残差, 并记为 $e1$ 。将残差与其一阶滞后的散点图与线性拟合图画在一起。根据此图, 是否可能存在自相关?
- (3) 画残差的自相关图。
- (4) 使用 Q 检验, 检验扰动项是否存在自相关。
- (5) 使用 HAC 标准误进行回归, 将截断参数设为 $n^{1/4}$ 。
- (6) 使用迭代式 CO 估计法进行 FGLS 估计。
- (7) 使用迭代式 PW 估计法进行 FGLS 估计。
- (8) 考虑到消费可能存在惯性, 将被解释变量 $lgasq$ 的一阶滞后作为解释变量, 加入回归方程。此滞后项是否显著?
- (9) 对于修改后的模型, 再次进行 Q 检验。是否还存在自相关?

5. 使用数据集 `acemoglu.dta` 复制 Acemoglu *et al.* (2001) 的部分结果。该数据集包含 64 个曾为欧洲殖民地的国家, 主要变量为 $\logpgp95$ (1995 年人均 GDP, 购买力平价), $avexpr$ (1985-1995 年间的平均产权保护程度, 0 为最低, 10 为最高), lat_abst (首都纬度的绝对值除以 90), 以及 $\logem4$ (殖民者死亡率的对数)。另外, 变量 `shortnam` 以三个字母作为每个国家的简称。

(1) 为了直观地考察产权保护与经济发展的关系, 将 $\logpgp95$ 与 $avexpr$ 的散点图与线性拟合图画在一起, 并为每个散点标注国家简称。

(2) 使用稳健标准误, 把 $\logpgp95$ 对 $avexpr$ 及 lat_abst 进行回归, 评论变量系数的符号、统计显著性及经济意义。

(3) 由于 $avexpr$ 可能为内生解释变量, 使用 $\logem4$ 作为 $avexpr$ 的工具变量, 重新进行 (2) 的回归。工具变量回归的结果与 OLS 有何不同?

(4) $\logem4$ 是否为弱工具变量?