

第 2 次作业

提交日期：4 月 14 日

1. 考虑一般的线性回归模型

$$Y_{N \times 1} = X_{N \times K} \beta_{K \times 1} + \varepsilon_{N \times 1},$$

其中 K 个解释变量 $X = [Z_{N \times J}, W_{N \times M}]$ 可以分为两组 Z 与 W , 满足 $J + M = K$; 同时 $\beta = [\delta^T, \theta^T]^T$ 也可以分为对应的两组。故回归模型可以等价的表示为:

$$Y_i = Z_i^T \delta + W_i^T \theta + \varepsilon_i \iff Y = Z\delta + W\theta + \varepsilon, \quad (1)$$

其中 ε 为误差向量。

(a) 定义矩阵 $P_X = X(X^T X)^{-1} X^T$ 。请验证 P_X 满足如下两条性质:

- i. 对于任意一个 X 的列向量线性组合构成的向量 $\xi = Xa$, $\forall a \in \mathbb{R}^K$, $P_X \xi = \xi$ 。
- ii. 对 \mathbb{R}^N 中任意向量 ζ , $\zeta - P_X \zeta = (I - P_X)\zeta$ 与 $\xi = Xa$ 相互垂直, $\forall a \in \mathbb{R}^K$, 即前者转置与后者乘积为 0。

这样的矩阵 P_X 称为关于 X (列线性空间) 的投影矩阵。

(b) 令 $\hat{\beta} = [\hat{\delta}^T, \hat{\theta}^T]^T$ 为回归模型 (1) 系数的 OLS 估计, $e = Y - X\hat{\beta}$ 。与 (a) 类似地定义 P_W 。令

$$\tilde{Y} = (I - P_W)Y, \quad \tilde{Z} = (I - P_W)Z.$$

请证明下述结论:

- i. \tilde{Y} 与 \tilde{Z} 分别为 Y 与 Z 对 W 回归的残差向量;
- ii. e 垂直于 Z 和 W , 即 $Z^T e$ 与 $W^T e$ 均为 0 向量;
- iii. e 在 W 上的投影为 0 , 即 $P_W e = 0$;
- iv. 考虑 \tilde{Y} 对 \tilde{Z} 的回归 $\tilde{Y} = \tilde{Z}\delta + u$, u 为新的误差向量, 请说明这个回归的 OLS 估计系数正好等于 (1) 中 OLS 估计的结果 $\hat{\delta}$ 。提示: 在 (1) 的 OLS 估计式 $Y = Z\hat{\delta} + W\hat{\theta} + e$ 两边同乘 $I - P_W$, 进而验证 $\hat{\delta}$ 满足 $\tilde{Y} = \tilde{Z}\delta + u$ 回归系数 OLS 估计的条件。

至此, 你已经证明了著名的 Frisch-Waugh-Lovell 定理, 即回归模型中控制变量 W_i^T 的加入, 意味着主解释变量 Z_i^T 对因变量 Y_i 的影响大小 (即回归系数估计值), 已经控制了 W_i^T 对因变量、主解释变量各自的影响, 具有“保持控制变量不变, 解释变量对因变量边际影响”的含义。

(c) 作为 FWL 定理的应用, 请证明: 在包含常数项的回归模型 $Y_i = \alpha + X_i^T \beta + e_i$, $i = 1, \dots, N$ 中, OLS 估计量 $\hat{\beta}$ 等于对所有变量去除均值后的回归估计值, 即对 $\tilde{Y}_i = \tilde{X}_i^T \beta + u_i$ 的 OLS 估计, 其中 $\tilde{Y}_i = Y_i - \bar{Y}$, \bar{Y} 为 Y_i 样本均值, \tilde{X} 定义类似。

- (d) 考虑如下数值示例：样本量 $N = 6$ ，回归变量个数 $K = 3$ ，回归方程为 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ，其中

$$\mathbf{Y}^T = \begin{bmatrix} 2 & 1 & 3 & 3 & 2 & 2 \end{bmatrix}, \quad (2)$$

$$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 \\ 3 & 2 & 2 & 1 & 0 & 1 \end{bmatrix}, \quad (3)$$

等价的，回归方程可写为 $\mathbf{Y} = \alpha \mathbf{1} + \delta \mathbf{Z} + \theta \mathbf{W} + \boldsymbol{\varepsilon}$ ，其中 $\mathbf{1}$ 为所有元素等于 1 的向量，即 \mathbf{X} 第 1 列， \mathbf{Z} 为 \mathbf{X} 第 2 列， \mathbf{W} 为 \mathbf{X} 第 3 列， $\boldsymbol{\beta} = [\alpha, \delta, \theta]^T$ 。

- i. 请列用 (2)–(3) 中的数值，计算回归系数向量 $\hat{\boldsymbol{\beta}} = [\hat{\alpha}, \hat{\delta}, \hat{\theta}]$ 的取值，以及回归的 R^2 。建议：利用 AI 工具，在 Python 或者 R 中进行编程计算。Stata 进行矩阵、向量编程计算的语法更繁琐，不建议使用。当然，上述矩阵计算仅涉及 3×3 矩阵求逆，手算亦可完成。
- ii. 基于 (2)–(3) 取值，验证上述 (c) 问的结论，即先用 Y_i, Z_i, W_i 分别对常数项 1 回归，再用所得残差 \tilde{Y}_i 对残差 \tilde{Z}_i, \tilde{W}_i 进行回归，并验证两个系数 $\tilde{\delta}, \tilde{\theta}$ 估计值与 i 中 $\hat{\delta}, \hat{\theta}$ 相等。请问此回归的 R^2 与 i 中 R^2 是否相等？是变高还是变低？
- iii. 定义 $\bar{\mathbf{W}} = [\mathbf{1}, \mathbf{W}]$ ，并定义 $\tilde{\mathbf{Y}}, \tilde{\mathbf{Z}}$ 为 \mathbf{Y}, \mathbf{Z} 各自对 $\bar{\mathbf{W}}$ 的回归残差，计算 $\tilde{\mathbf{Y}}$ 对 $\tilde{\mathbf{Z}}$ 的回归系数 $\tilde{\delta}$ ，并验证其等于 i 中估计值 $\hat{\delta}$ ，进而验证 (b) 中 iv，即 FWL 的一般结论。
- iv. 接续问题 i，计算 $\hat{\alpha}, \hat{\delta}, \hat{\theta}$ 的标准误， t -统计量取值，并对 3 个单系数原假设 $H_0 : \alpha = 0, H_0 : \delta = 0, H_0 : \theta = 0$ 分别计算 p -值，并在 5% 显著性水平下，判断 3 个原假设各自是否成立。建议：Python, R 或 Stata 中，都有标准的回归命令，可以直接输出上述计算结果；类似的，请使用 AI 工具学习如何编写简单代码，完成上述任务。
- v. 与 iv 类似，检验联合系数假设 $H_0 : \alpha = \delta = \theta = 0$ ，即计算回归模型的 F 统计量取值，计算相应的 p -值，并说明在 5% 显著性水平下，是否拒绝原假设？换言之，此时 F 统计量拒绝域临界值是多少？注意，首先确定此时 F -统计量所满足的 $F(\ell, m)$ 分布中的两个自由度 ℓ, m 的取值。