高级计量经济学

第三次作业

邓皓天　2023310114

1、 对于回归模型$Y_i = \beta_0 + \beta_1 X_i + \mu_i$，如果随机解释变量$X$与随机误差项$\mu$相关，现存在$X$的一个工具变量$Z$。请求出采用工具变量法时（以$Z$作为工具变量）有关待估参数$\beta_1$的方差。

答： 以Z作为工具变量得到的系数β_1的估计结果为

$$\beta_1 = \frac{\sum z_i y_i}{\sum z_i x_i} = \frac{\sum z_i (Y_i - \bar{Y})}{\sum z_i x_i} = \frac{\sum z_i Y_i}{\sum z_i x_i} - \frac{\bar{Y} \sum z_i}{\sum z_i x_i} = \frac{\sum z_i Y_i}{\sum z_i x_i}$$

因此其方差为

$$\begin{aligned}\mathrm{Var}(\beta_1) &= \mathrm{Var}(\frac{\sum z_i Y_i}{\sum z_i x_i} \mid X) = \sum (\frac{z_i}{\sum z_i x_i})^2 \mathrm{Var}\left[(\beta_0 + \beta_1 X + \mu) \mid X\right] \\ &= \sum (\frac{z_i}{\sum z_i x_i})^2 \cdot \sigma_\mu^2 = \frac{\sum z_i^2}{(\sum z_i x_i)^2} \cdot \sigma_\mu^2\end{aligned}$$

又因为$\sum x_i^2 = n\sigma_X^2$，$\sum z_i^2 = n\sigma_Z^2$,$(\sum z_i x_i)^2 = n^2 \rho_{X,Z}^2 \cdot \sigma_X^2 \cdot \sigma_Z^2$。所以

$$\mathrm{Var}(\beta_1) = \frac{\sigma_\mu^2}{n\sigma_X^2 \rho_{X,Z}^2}$$

2、 对于符合古典假设的回归模型$Y_i = \beta_0 + \beta_1 X_i + \mu_i$而言，如果不存在$X$与随机误差项$\mu$相关，那么可以采用OLS对于参数$\beta_1$进行估计。如果同时在经济学理论上能够找到另外一个变量Z作为X的工具变量，那么也可以采用工具变量法对于参数β_1进行估计。试问，在以上情况下，即既可以采用OLS估计β_1，也可以采用IV（工具变量法）估计β_1的情况下，你认为该选择哪种方法来进行估计参数更加有优势？为什么？请加以证明。

答： 如果模型满足古典假设，即误差项μ与解释变量X是独立的，那么OLS估计量是最优的，因为在这种情况下，OLS估计量是BLUE（最佳线性无偏估计量）。这意味着在所有无偏的线性估计量中（包括IV估计的β_1），OLS估计量具有最小的方差。由于$\rho^2_{X,Z} \in [0,1]$，因此

$$\mathrm{Var}(\beta_1^{\mathrm{OLS}}) = \frac{\sigma_\mu^2}{n\sigma_X^2} \leqslant \frac{\sigma_\mu^2}{n\sigma_X^2\rho_{X,Z}^2} = \mathrm{Var}(\beta_1^{\mathrm{IV}})$$

3、 对于回归模型$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \mu_i$，如果随机解释变量$X_2$与随机误差项$\mu$相关，请回答：

（1）现找到X_2的一个工具变量Z。对于以上模型可以分别采用2SLS和IV法来进行参数估计，请分别写出参数估计量的表达式。在此基础上，试问2SLS和IV法所估计参数是等价的吗？为什么？

（2）如果现在可以给X_2找到两个工具变量Z_1和Z_2，那么能够同时使用到两个工具变量的信息并采用IV法来对以上模型相关参数进行估计吗？为什么？如果不能，假设这时候可以选用任何一个工具变量的信息（采用IV法）来进行参数估计，请问IV所得结果与2SLS估计结果等价吗？为什么？请详细说明。

（3）通过以上（1）和（2）的分析，你对于2SLS和IV法的相互关系有什么新发现？

答： （1）此时2SLS和IV法所估计参数是等价的。

2SLS：首先对模型$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \mu_i$进行第一阶段的OLS估计，得到

$$\hat{X}_{2,i} = \gamma_0 + \gamma_1 X_{1,i} + \gamma_2 Z_i + \varepsilon_i$$

用$\hat{X}_{2,i}$替换$X_{2,i}$得

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 \hat{X}_{2,i} + \mu_i'$$

对其进行第二阶段的OLS估计即可得到$\hat{\beta}_2^{\mathrm{2SLS}}$。

IV法可直接得倒参数估计量

$$\hat{\beta}_2^{\mathrm{IV}} = \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(X_{2,i}, Z_i)}$$

此时2SLS和IV法所估计的参数是等价的。因为在2SLS的第二阶段，我们实际上是在用Z_i作为$X_{2,i}$的工具变量来进行估计，这与IV法的思想是一致的。

（2）可以将Z_1和Z_2合并为一个工具变量Z，然后采用Z来进行IV估计，即可同时使用到两个工具变量的信息。这种"合成"方法并不一定是2SLS中第一阶段的OLS，还可以是其他方法（例如机器学习等）。如果"合成"方法为OLS，则此时IV法与2SLS等价。

（3）2SLS可以看作是IV法的一个扩展，可以处理多个工具变量或者多个内生解释变量（过度识别）。而IV法则是更为基础的方法，它主要用于处理单个内生解释变量和单个工具变量（恰好识别）的情况。