

PS 5 - Instrumental Variables

1. Estimador de Wald

Suponga un modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

donde x_i es potencialmente endógena. Además, suponga que el instrumento, z_i es una variable binaria. Muestre que el estimador IV en este caso es

$$\beta_1^{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

donde \bar{y}_1, \bar{x}_1 (\bar{y}_0, \bar{x}_0) representan las medias cuando $z = 1$ ($z = 0$).

Suponiamos que la asignación al tratamiento es aleatoria.

¿Qué pasa si vemos endogeneidad? (Selección en inobservables)

| | | | |
|---------------|---------------|--|---------------------|
| Always takers | \rightarrow | $D=1, D=0$ | \Rightarrow $T=1$ |
| Never takers | \rightarrow | $D=1, D=0$ | \Rightarrow $T=0$ |
| Compliers | \rightarrow | $D=1 \Rightarrow T=1, D=0 \Rightarrow T=0$ | |
| Defiers | \rightarrow | $D=1 \Rightarrow T=0, D=0 \Rightarrow T=1$ | |

IV \rightarrow Identifica LATE (Local Average Treatment Effects) sobre compliers.

Necesito:

- Relevancia $cov(x, z) \neq 0$
- Exogeneidad $cov(z, u) = 0$
- Exclusión : z no aparece en la variable de interés.

2. Estimador de Wald con Datos Simulados

2. Estimador de Wald con datos simulados

En este ejercicio se propone extender la simulación del Problem Set 1 a un marco en el que la asignación del tratamiento y quienes resultan tratados no son iguales.

1. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.

2. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, \dots, n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal.
3. Genere una variable aleatoria uniforme en el intervalo $[0, 1]$. Con ella, genere variables que indiquen el tipo de individuo. Utilice: *always taker* si la variable es menor a 0.25, *never taker* si la variable está entre 0.25 y 0.5, *defier* si la variable está entre 0.5 y 0.75 y *complier* si la variable es mayor a 0.75. Genere la variable de si los individuos toman el tratamiento o no dependiendo del grupo en el que están.
4. Genere la variable Y observada como $Y = DY_1 + (1 - D)Y_0$.
5. Estime el LATE y compare con el ATE.

