Microeconometría II

Maestrías en Economía y Econometría

Examen Final
Segundo Trimestre 2025

Propiedades de Muestra Finita de los Estimadores de Impacto de una Política

El trabajo requiere el uso de simulaciones de Monte Carlo para investigar las propiedades de muestra finita de distintos estimadores del efecto de un tratamiento. Consiste en simular datos bajo procesos de generación de datos (DGP) conocidos y evaluar los estimadores para determinar sesgo, varianza, error medio cuadrático y tasas de cobertura de los intervalos de confianza.

Asuma un tratamiento binario D_i (1 = tratado, 0 = control) y una variable de resultados Y_i. El verdadero efecto promedio del tratamiento (ATE) se conoce en cada simulación. Utilice 1,000 replicaciones de Monte Carlo en cada ejercicio, con muestras de tamaño pequeño $N = 100$ o $N = 200$ (como se especifique). Reporte sus resultados en un informe que incluya tablas con medias y desvíos estándar a traves de las simulacioness.

Ejercicio 1: Diferencia de medias y ajuste por regresión

Objectivo: comparar diferencias en medias simples (sin ajustar) con estimadores de ajuste por regresión bajo selección sobre observables.

DGP:

- Genere la variable observable: $X_i \sim N(0, 1)$.
- Asignación del tratamientot: $D_i = 1$ si $X_i + \epsilon_i > 0$, 0 en otro caso, donde $\epsilon_i \sim N(0, 1)$ (sesgo de selección vía X_i).
- Variable de resultado: $Y_i = 2 + 3D_i + 2X_i + u_i$, donde $u_i \sim N(0, 1)$. Valor verdadero: ATE = 3.

Para ($N = 100$ y $N = 200$):

1. Estime el ATE usando diferencia en medias: $\hat{\tau}_{DM} = \bar{Y}_{D=1} - \bar{Y}_{D=0}$.
2. Estime el ATE usando MCC (OLS): $Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + e_i$, donde $\hat{\tau}_{OLS} = \hat{\beta}_1$.
3. Cuál es el supuesto implícito aquí para que $\hat{\tau}_{OLS}$ identifique el ATE?

4. Para cada estimador, calcule a traves de las 1,000 simulaciones: sesgo (media $\hat{\tau} - 3$), varianza, error medio cuadrático (MSE), y tasa de cobertura de intervalos de 95% de confianza (IC) (proporción of ICs que contienen al verdadero ATE = 3).
5. Discuta: Cómo la inclusión de X_i reduce el sesgo? Cómo mejoran las propiedades con un N más grande?

Ejercicio 2: Propensity Score Matching

Objetivo: Examinar el rendimiento de los estimadores de emparejamiento bajo independencia condicional.

DGP:

- Genere dos variables: $X_{1i} \sim N(0,1)$, $X_{2i} \sim Bernoulli(0.5)$.
- Propensity score: $P(D_i = 1|X) = \frac{1}{1+\exp(-(0.5+X_{1i}+2X_{2i}))}$. Extraiga D_i desde una Bernoulli con esta probabilidad.
- Variable de resultado: $Y_i = 1 + 4D_i + X_{1i} + 3X_{2i} + u_i$, $u_i \sim N(0,1)$. Valor verdadero: ATE = 4.

Para ($N = 100$ y $N = 200$):

1. Estime el propensity score usando un Logit: $\hat{P}(D_i = 1|X_1, X_2)$.
2. Use emparejamiento del vecino más cercano (nearest-neighbor matching, 1:1, con reemplazo) para estimar el ATE: promedio de las diferencias emparejadas $Y_i(D = 1) - Y_j(D = 0)$.
3. Calcule a traves de las 1,000 simulaciones: sesgo, varianza, error medio cuadrático (MSE), y tasa de cobertura de intervalos de 95% de confianza (IC) (use bootstrap para estimar los errores estándar, 200 muestras por réplica).
4. Discuta: Cómo afecta el desbalance de las características observables al sesgo en muestras pequeñas? Compare con lo que obtuvo en el Ejercicio 1.

Ejercicio 3: Variables Instrumentales (IV)

Objetivo: Simular problemas de instrumentos débiles y mínimos cuadrados en dos etapas (MC2E) en muestras pequeñas.

DGP:

- Instrumento $Z_i \sim Bernoulli(0.5)$.
- Tratamiento: $D_i = 0.2 + 0.3Z_i + 0.5X_i + v_i$, $v_i \sim N(0,1)$ (endogeneidad via errores correlacionados).
- Variable de resultado: $Y_i = 5 + 2D_i + X_i + u_i$, donde $u_i = 0.8v_i + \epsilon_i$, $\epsilon_i \sim N(0,1)$, $X_i \sim N(0,1)$. Valor verdadero: ATE = 2.

- Para una variante de instrumentos débiles: Reduzca la fuerza del instrumento a $D_i = 0.2 + 0.05Z_i + 0.5X_i + v_i$.

Para ($N = 100$):

1. Estime via "naive" MCC (sesgado debido a la endogeneidad).
2. Estime via MC2E: Primera etapa: D_i sobre Z_i, X_i; segunda etapa Y_i sobre $\hat{D}_i, X_i$.
3. Para instrumentos fuertes y débiles: Calcule sesgo, varianza, MSE, cobertura, y el estadístico-F de la primera etapa.
4. Discuta: Cómo amplifica el instrumento débil el sesgo/varianza en muestra pequeña? Relaciónelo con las propiedades de muestra finita de IV.

Ejercicio 4: Diferencia-en-Diferencias (DID)

Objetivo: Evaluar violaciones del supuesto de tendencias paralelas en muestras pequeñas.

DGP (Panel con 2 períodos, pre/post):

- Unidades $i = 1, ..., N$, tiempo $t = 1, 2$.
- Tratamiento: mitad de las unidades son tratadas en $t = 2$ (group fijo).
- Variable observada: $X_{it} \sim N(0, 1)$.
- Variable de resultado: $Y_{it} = \alpha_i + \lambda_t + 1.5D_{it} + 2X_{it} + u_{it}$, $u_{it} \sim N(0, 1)$, donde $D_{it} = 1$ si está en el grupo de tratamiento y $t = 2$. Valor verdadero: ATT $= 1.5$.
- Variante que viola tendencias paralelas: agregue pre-tendencia para el tratamiento: $\lambda_t = 0.5t$, y 0 para el control.

Para ($N = 100$ unidades):

1. Estime DID: $\hat{\tau}_{DID} = (\bar{Y}_{treated,post} - \bar{Y}_{treated,pre}) - (\bar{Y}_{control,post} - \bar{Y}_{control,pre})$.
2. Incluya en la regresión heterogeneidad de corte transversal y temporal y X_{it}.
3. Para el caso base y el que viola el supuesto de tendencias paralelas: Calcule sesgo, varianza, MSE, cobertura (use errores estándar por clúster al nivel de unidad de corte transversal -cluster SEs at unit level-.
4. Discuta: Cómo sesga la violación de tendencias paralelas las estimaciones en muestra pequeña? Compare con (N=500) más grande (extensión opcional).

Ejercicio 5: Síntesis y Comparación

1. Tabule los resultados de todos los estimadores de los 4 ejercicios analizados (sesgo, MSE, etc.) para el caso de $N = 100$.
2. Cuál de los estimadores piensa que funciona mejor/peor en muestras pequeñas? Cuándo podrían simulaciones como estas influir/orientar decisiones empíricas reales?

Instrucciones para la entrega:

- El trabajo es individual.
- Use los últimos 5 números de su documento de identidad como seed en el código para que los resultados sean replicables.
- El trabajo se puede hacer en Stata, R, Matlab o similar.
- Hay que entregar un archivo .pdf (.doc o similar) con un reporte de sus respuestas y conclusiones y adjuntar el código (.do, .r, .m, etc.) o entregar el .log file aparte.
- En el código deberán hacer comentarios breves para que se entienda el procedimiento.
- La entrega se realizará vía Campus Virtual hasta el día lunes 17 de noviembre de 2025 a las 11.59pm.