

**Microeconometría II**  
**Examen Final**

---

Considere un programa de capacitación laboral dirigido a jóvenes con poca experiencia en el mercado de trabajo. Estamos interesados en estimar el impacto promedio del programa sobre el ingreso laboral ( $y$ ) de la población de referencia (ATE) y sobre el ingreso laboral de los beneficiarios del programa (ATT). Para esta evaluación se le pide comparar la estimación del impacto usando los métodos de regresión y de emparejamiento. Para esto genere 5000 observaciones de las siguientes variables:

- Antes de generar los datos utilice como semilla el valor 7672.
- *Edad* se genera a partir de una distribución normal. Para el grupo de tratamiento  $Edad \sim N(20; 2,5)$  y para el grupo de control  $Edad \sim N(25; 3)$ .
- *Experiencia* en el mercado de trabajo (en años) se genera a partir de una distribución normal. Para el grupo de tratamiento  $Experiencia \sim N(1,76; 0,5)$  y para el grupo de control  $Experiencia \sim N(2,3; 0,75)$ .
- Construya estas dos variables centradas con respecto a sus medias ( $Edad_c$ ,  $Experiencia_c$ ) y luego calcule las variables al cuadrado ( $Edad_c^2$ ,  $Experiencia_c^2$ ) y el producto cruzado ( $Edad_c \times Experiencia_c$ ).
- $\epsilon \sim N(0; 0,5)$
- *tratamiento* es una variable binaria que vale 0 para las observaciones 1 a 2500 y vale 1 para las observaciones 2501 a 5000.
- *Resultados potenciales*  $y_0$  es el resultado potencial sin tratamiento e  $y_1$  es el resultado potencial con tratamiento:

$$\begin{aligned}y_0 &= 15000 + 10,25 \times Edad_c + 10,5 \times Edad_c^2 + 1000 \times Experiencia_c \\ &\quad - 10,5 \times Experiencia_c^2 + 500 \times Edad_c \times Experiencia_c + \epsilon \\ y_1 &= y_0 + 2500 + 100 \times Edad_c + 1000 \times Experiencia_c\end{aligned}$$

- Calcule y reporte el verdadero ATE y el verdadero ATT a partir de esta generación de datos.
1. Ahora cambie la semilla (utilice como valor los últimos tres números de su registro) y realice 1000 simulaciones con la misma especificación pero en lugar de reportar para cada simulación el verdadero efecto promedio del programa (ATE) estímelo a partir de un modelo de regresión con las siguientes especificaciones:

a) Efecto tratamiento homogéneo y regresión sin términos cuadráticos

$$y = \beta_0 + \beta_1 \text{tratamiento} + \beta_2 Edad_c + \beta_3 Experiencia_c + \varepsilon \quad (1)$$

b) Efecto tratamiento homogéneo y regresión con términos cuadráticos e interacción

$$y = \beta_0 + \beta_1 \text{tratamiento} + \beta_2 \text{Edad}_c + \beta_3 \text{Edad}_c^2 + \beta_4 \text{Experiencia}_c + \beta_5 \text{Experiencia}_c^2 + \beta_6 \text{Edad}_c \times \text{Experiencia}_c + \varepsilon \quad (2)$$

Reporte el promedio de la estimación del ATE en las 1000 simulaciones para estos dos casos. Muestre un gráfico con la función de densidad del estimador para cada caso. En el gráfico marque el valor estimado promedio y el verdadero valor del ATE. Que conclusiones puede sacar de la estimación del impacto del programa sobre la población de referencia utilizando ambos métodos? Alguno de los estimadores es “mejor” que el otro? discuta los resultados.

2. Usando las mismas 1000 simulaciones cambie las especificaciones de las regresiones por modelos saturados:

a) Saturación simple

$$y = \beta_0 + \beta_1 \text{tratamiento} + \beta_2 \text{Edad}_c + \beta_3 \text{Experiencia}_c + \beta_4 (\text{tratamiento} \times \text{Edad}_c) + \beta_5 (\text{tratamiento} \times \text{Experiencia}_c) + \beta_6 \text{Edad}_c \times \text{Experiencia}_c + \varepsilon \quad (3)$$

b) Saturación completa

$$\begin{aligned} y = & \beta_0 + \beta_1 \text{tratamiento} + \beta_2 \text{Edad}_c + \beta_3 \text{Edad}_c^2 + \beta_4 \text{Experiencia}_c + \beta_5 \text{Experiencia}_c^2 \\ & + \beta_6 (\text{Edad}_c^2 \times \text{Edad}_c) + \beta_7 \text{Edad}_c \times \text{Experiencia}_c + \beta_8 (\text{Edad}_c \times \text{Experiencia}_c^2) \\ & + \beta_9 (\text{Edad}_c^2 \times \text{Experiencia}_c) + \beta_{10} (\text{Edad}_c^2 \times \text{Experiencia}_c^2) \\ & + \beta_{11} (\text{Experiencia}_c \times \text{Experiencia}_c^2) + \beta_{12} (\text{Edad}_c^2 \times \text{Edad}_c \times \text{Experiencia}_c) \\ & + \beta_{13} (\text{Edad}_c^2 \times \text{Edad}_c \times \text{Experiencia}_c^2) + \beta_{14} (\text{tratamiento} \times \text{Edad}_c) \\ & + \beta_{15} (\text{tratamiento} \times \text{Edad}_c^2) + \beta_{16} (\text{tratamiento} \times \text{Experiencia}_c) \\ & + \beta_{17} (\text{tratamiento} \times \text{Experiencia}_c^2) + \beta_{18} (\text{tratamiento} \times \text{Edad}_c \times \text{Edad}_c) \\ & + \beta_{19} (\text{tratamiento} \times \text{Edad}_c \times \text{Experiencia}_c) + \beta_{20} (\text{tratamiento} \times \text{Edad}_c \times \text{Experiencia}_c^2) \\ & + \beta_{21} (\text{tratamiento} \times \text{Edad}_c^2 \times \text{Experiencia}_c) + \beta_{22} (\text{tratamiento} \times \text{Edad}_c^2 \times \text{Experiencia}_c^2) \\ & + \beta_{23} (\text{tratamiento} \times \text{Experiencia}_c \times \text{Experiencia}_c^2) \\ & + \beta_{24} (\text{tratamiento} \times \text{Edad}_c^2 \times \text{Edad}_c \times \text{Experiencia}_c) \\ & + \beta_{25} (\text{tratamiento} \times \text{Edad}_c^2 \times \text{Edad}_c \times \text{Experiencia}_c^2) + \varepsilon \end{aligned} \quad (4)$$

Reporte el promedio de la estimación del ATE en las 1000 simulaciones para estos dos casos. Muestre un gráfico con la función de densidad del estimador para cada caso. En el gráfico marque el valor estimado promedio y el verdadero valor del ATE. Que conclusiones puede sacar de la estimación del impacto del programa sobre la población de referencia utilizando ambos métodos? Compare estos resultados con los del punto anterior. Que podemos aprender de este ejercicio? Discuta los resultados.

3. Repita los dos puntos anteriores pero ahora estime el ATT.

4. En este punto se le pide estimar el ATT utilizando el método de emparejamiento (matching) visto en clase. Utilice los estimadores de Abadie e Imbens. La estimación que se le pide usa cuatro modelos diferentes de emparejamiento. Para cada modelo, minimice la distancia de Mahalanobis mencionada en clase para obtener las mejores coincidencias posibles (el emparejamiento es uno a uno).

- *Modelo 1:* emparejamiento usando *Edad* y *Experiencia* solamente y use el estimador de Abadie e Imbens sin corregir por sesgo.

- *Modelo 2*: emparejamiento usando *Edad* y *Experiencia* solamente y use el estimador de Abadie e Imbens que corrige por sesgo.
  - *Modelo 3*: emparejamiento usando *Edad* y su cuadrado *Experiencia* y su cuadrado y la interacción entre *Edad* y *Experiencia* y use el estimador de Abadie e Imbens sin corregir por sesgo.
  - *Modelo 4*: emparejamiento usando *Edad* y su cuadrado *Experiencia* y su cuadrado y la interacción entre *Edad* y *Experiencia* y use el estimador de Abadie e Imbens que corrige por sesgo.
- a) Para cada una de las 1000 simulaciones estime el ATT como se detalla en cada modelo y reporte el promedio del ATT para los cuatro casos. Muestre un gráfico con la densidad del estimador en cada uno de los 4 casos, el valor estimado promedio y el verdadero valor del ATT. Qué conclusiones puede sacar de la estimación del impacto del programa sobre la población beneficiaria utilizando estos métodos de emparejamiento? Compare estos resultados con las estimaciones del ATT usando el método de regresión. Que podemos aprender de este ejercicio? Describa las conclusiones de su análisis.

### Instrucciones para la entrega:

- El trabajo es **individual**.
- El trabajo se puede hacer en Stata, R, Python, Matlab, etc.
- Hay que entregar un archivo .pdf con un reporte que responda todas las preguntas del examen y adjuntar el código (.do, .r, o .m) o entregar el .log file aparte.
- Los archivos deberán tener el apellido de quien entregue y la seed utilizada en las simulaciones, es decir, si Caravello usó el seed 668 los archivos se llamarán **Caravello668.pdf**, **Caravello668.do**, **Caravello668.log**, **Caravello668.m** o **Caravello668.r**
- En el código deberán hacer **comentarios breves** para que se entienda el **procedimiento**.
- La entrega se realizará **vía Campus Virtual** hasta el día **lunes 23 de octubre de 2023 a las 11.59pm**.