

---

**IX XORNADA DE  
USUARIOS DE  
EN GALICIA** 

Santiago de Compostela, 20 de outubro de 2022

---

Matching para o estudo  
do impacto de políticas  
públicas

**José Manuel Amoedo**

Departamento de Economía  
Aplicada. Grupo de Investigación  
ICEDE, Facultade de Ciencias  
Económicas e Empresariais da USC

---

# Contidos

Introdución

Matching para a análise de políticas públicas: o caso da CPI (Compra Pública de Innovación)

Conclusións

# Introducción

- O matching tamén é coñecido como emparellamento ou pareo.
  - O obxectivo do matching é igualar, ou equilibrar, a distribución das covariables entre os grupos de tratamento e control para converter unha mostra non aleatoria nunha mostra coas características dunha mostra aleatoria.
  - Os usos do matching abordan principalmente a estimación de efectos causais, pero tamén efectos non causais.
-

---

# Usos do matching

## Efectos causais

### Efecto do tabaco na esperanza de vida

- Precisamos un grupo de tratamento (fumadores) e un grupo de control (non fumadores).
- Precisamos que ambos grupos sexan similares (idade, hábitos de vida, patoloxías previas...).
- O matching leva a que ambos grupos presenten características (covariables) similares.
- A diferenza na idade media entre ambos grupos equilibrados mide o efecto do tabaco na esperanza de vida.

## Efectos non causais

### Desigualdades de xénero

- Precisamos un grupo de tratamento (mulleres) e un grupo de control (homes).
  - Precisamos que ambos grupos sexan similares (idade, formación, sector laboral, estrutura familiar...).
  - O matching leva a que ambos grupos presenten características (covariables) similares.
  - A diferenza no nivel de renda entre ambos grupos equilibrados mide a desigualdade entre ambos sexos no nivel de renda.
-

# Matching para a análise de políticas públicas: procedemento

Os pasos a seguir á hora de levar a cabo o matching son os seguintes:

1. Identificar a variable de tratamento (neste caso que a empresa tivera un contrato de CPI).
2. Seleccionar as covariables a incluír no modelo (nº traballadores, facturación, gasto en I+D, sector...).
3. Seleccionar a medida da distancia a empregar (Propensity Score).
4. Seleccionar a metodoloxía a empregar para o matching (Nearest neighbor).
5. Recursos para o seu uso en R (A librería MatchIt).

# Matching para a análise de políticas públicas: procedemento (1)

Identificar a variable de tratamento (neste caso que a empresa tivera un contrato de CPI).

- a. Dita variable debe ser unha binomial (1 se tivo CPI e 0 se non a tivo).
  - b. Debemos ter en conta outros posibles tratamentos que entren en conflito coa nosa (subvencións de apoio á I+D).
  - c. Analizar se é posible excluír ás empresas con apoio á I+D para ter como grupo de tratamento empresas **SO** con CPI e como control empresas sen CPI e sen apoio á I+D.
-

# Matching para a análise de políticas públicas: procedemento(2)

Seleccionar as covariables a incluír no modelo.

- a. Debemos incluír todas as covariables relevantes á hora de explicar porque unha empresa ten CPI ou non con anterioridade ó tratamento (mellor de máis que de menos).
  - b. Se hai tratamentos alternativos (apoio á I+D) que non se poden excluír debemos incluír unha covariable binomial que o recolla.
  - c. É importante basearse tanto nos resultados obtidos nas seguintes fases como na literatura existente á hora de valorar que covariables son relevantes.
-

# Matching para a análise de políticas públicas: procedemento(3)

Seleccionar a medida da distancia a empregar.

- A distancia  $D_{ij}$  pode ser definida, para dos individuos calquera como unha medición da similitude entre ambos individuos  $i$  e  $j$ .
- As máis empregadas son a exacta, a de Mahalanobis, a Propensity Score (PS) e a Linear Propensity score.
- A PS ( $e_j$ ) raramente é coñecida e debe ser estimada. Serve calquera modelo que relacione unha variable binaria (CPI) cunha serie de regresores (covariables).
- No modelo os aspectos nos que centrarse non son os da econometría común. So serve para medir a similitude dos individuos.

Exacta

$$D_{ij} = \begin{cases} 0 & \text{if } X_i = X_j \\ \infty & \text{if } X_i \neq X_j, \end{cases}$$

Mahalanobis

$$D_{ij} = (X_i - X_j)' \sum^{-1} (X_i - X_j)$$

Propensity Score

$$D_{ij} = |e_i - e_j|,$$

Linear Propensity score

$$D_{ij} = |\text{logit}(e_i) - \text{logit}(e_j)|,$$



# Matching para a análise de políticas públicas: procedemento(4)

Seleccionar a metodoloxía a empregar para o matching. Dous tipoloxías:

- a. Nearest neighbor matching.
  - i. 1:1 matching.
  - ii. Optimal matching.
  - iii. Ratio matching (k:1).
    - 1. Caliper.
    - 2. Reempresamento.

b. Subclassification, full matching e weighting.

- i. Empregan ponderadores e inclúen a todos os individuos de control.
- ii. Subclassification: crea subclases con individuos similares.
- iii. Full matching: selecciona o nº de subclases automáticamente.
- iv. Weighting: usa a PS para ponderar os individuos en ambos grupos.

# Matching para a análise de políticas públicas: recursos en R

- As librerías máis coñecidas son MatchIt, Matching, twang, cem, optmatch, PSAgraphics, Synth, Cobalt e ebal (Stuart, E. A., 2010).
- A librería MatchIt (Ho, D. E., et al., 2011) permite empregar múltiples metodoloxías e distancias. Tamén permite obter diferentes medidas e gráficos para comprobar o equilibrio dos grupos, así como obter datasets para ambos grupos ou para o conxunto.
- Tamén presenta diferentes funcionalidades adicionais como “discard”, “reestimate”, “m.order”, “replace”, “ratio”, “caliper” e “exact”.

# Matching para a análise de políticas públicas: o caso da CPI en Colombia

- Orde do matching:

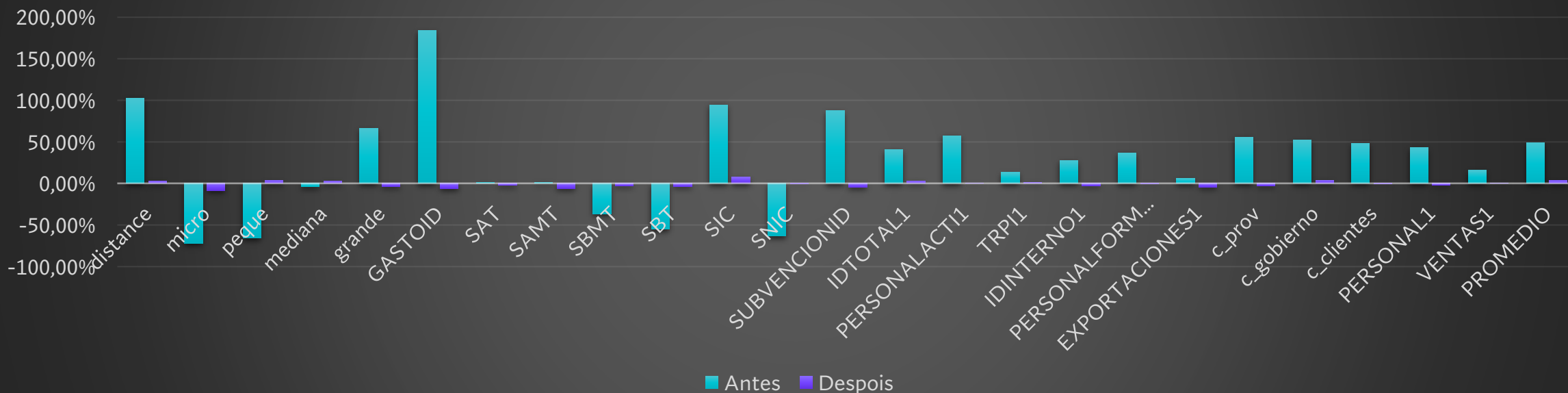
```
> equation <-CPI ~ micro +peque +mediana +grande + GASTOID +SAT + SAMT + SBMT +SBT +SIC+ SNIC  
+SUBVENCIONID + IDTOTAL1 + PACTI +TRPI + IDINTERNO1 + PFOR1 + EXP + c_prov + c_gobierno + c_clientes
```

```
> m.out <-matchit(formula = equation, data, method = "nearest", ratio = 50, distance = "probit", reestimate = TRUE,  
verbose = TRUE, replace = TRUE, discard = "both")
```

- Obtemos as medidas de equilibrio empregadas por Guerzoni e Raiteri (2015):
  - Sesgos medios estandarizados, pseudo- $R^2$ , test de verosimilitude e análise gráfica.
-

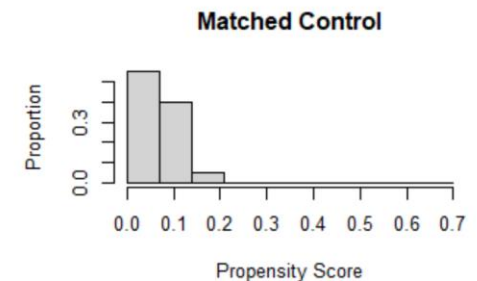
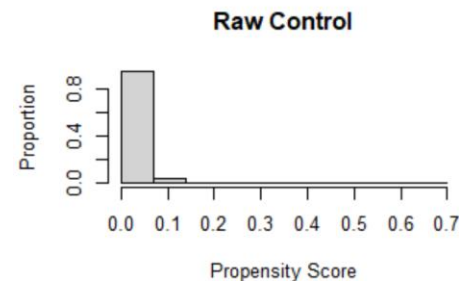
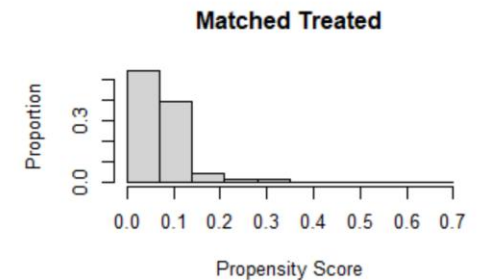
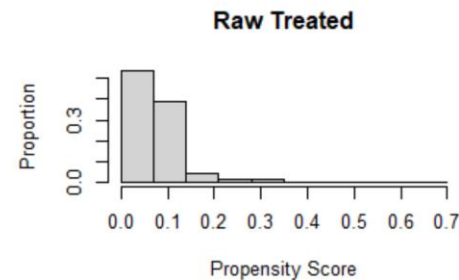
# Matching para a análise de políticas públicas: o caso da CPI en Colombia (1)

Sesgos medio estandarizados



# Matching para a análise de políticas públicas: o caso da CPI en Colombia (2)

|                        | Antes      | Despois |
|------------------------|------------|---------|
| Individuos tratados    | 450        | 450     |
| Individuos control     | 16.089     | 4.382   |
| Pseudo-R <sup>2</sup>  | 0,328      | 0,004   |
| Test de verosimilitude | 1352,85*** | 19,43   |



## Matching para a análise de políticas públicas: o caso da CPI en Colombia (3)

| Variable  | Medias ponderadas |            | Impacto (diferencia %) | Desviación estándar |           | Test-t/Test-u |
|---|-------------------|------------|------------------------|---------------------|-----------|---------------|
|   | Tratados          | Control    |                        | Tratados            | Control   |               |
| Nuevos bienes y servicios                             | 3,8533            | 1,5596     | 147,1%                 | 24,41               | 5,03      | 23,3732***    |
| Bienes y servicios mejorados                          | 1,7422            | 0,7523     | 131,6%                 | 3,37                | 2,50      | 1,8063***     |
| Nuevos métodos de prestación                          | 0,8400            | 0,6919     | 21,4%                  | 2,01                | 1,42      | 1,9841***     |
| Nuevos métodos organizativos                          | 0,7556            | 0,5749     | 31,4%                  | 1,94                | 1,18      | 2,6797***     |
| Nuevas técnicas de comercialización                   | 0,3978            | 0,3463     | 14,9%                  | 0,96                | 0,81      | 1,4100***     |
| Personal  | 621,28            | 638,93     | -2,8%                  | 1080,39             | 1163,37   | 0,8579***     |
| Ventas (millones pesos colombianos)                   | 229106,224        | 229244,563 | -0,1%                  | 1129827,36          | 975676,90 | 1,3330***     |
| Peso exportaciones                                    | 0,0361            | 0,0428     | -15,5%                 | 0,11                | 0,12      | 0,7420***     |
| Peso gasto I+D total sobre ventas                     | 0,0105            | 0,0134     | -21,8%                 | 0,04                | 0,08      | 0,2632***     |
| Peso gasto I+D interno sobre ventas                   | 0,0360            | 0,0302     | 19,3%                  | 0,08                | 0,10      | 0,6388***     |
| Peso personal formado sobre personal (1)              | 0,0602            | 0,0618     | -2,6%                  | 0,13                | 0,13      | 1,1294***     |
| Peso personal I+D sobre total                         | 0,0919            | 0,0804     | 14,3%                  | 0,12                | 0,12      | 0,9694***     |
| Patentes obtenidas                                    | 2,7578            | 3,6629     | -24,7%                 | 12,08               | 16,72     | 0,5188***     |
| (1) Significa a dicha variable se le aplica el test-u |                   |            |                        |                     |           |               |
| * p<0,1, **p<0,05, *** p<0,01                         |                   |            |                        |                     |           |               |

# Conclusións

- A CPI parece ter efectos significativos nas variables de resultados analizadas dunha forma similar ó que recolle a literatura existente.
- O matching amósase como unha boa metodoloxía para estudar estes aspectos.
- Ademais R, e máis concretamente a librería MatchIt, amósanse como ferramentas útiles para aplicar a metodoloxía en cuestión múltiples como as políticas públicas e outros aspectos nos que os estudos experimentais non se poden levar a cabo.

# Referencias

Stuart, E. A. (2010). Matching methods for casual inference: A review and a look forward. *Statistical Science*, 25(1), 1-21. <https://doi:10.2307/41058994>

Ho, D. E., Imai, K., King, G. e Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8) 1-28, <https://doi.org/10.18637/jss.v042.i08>

Guerzoni, M. e Raiteri, E. (2015). Demand-side vs. supply-side technology policies: Hidden treatment and new empirical evidence on the policy mix. *Research Polic*



---

**Moitas grazas pola vosa atención**

**Correo electrónico: [jm.amoedo@usc.es](mailto:jm.amoedo@usc.es)**

---