

Variable Instrumental

Matías Javier Deneken & María Paz Mujica

5 de junio, 2023

Preparativos

```
# libraries
pacman::p_load(foreign, dplyr, texreg, ggplot2, estimatr, stargazer, plm, car, panelr,
               clubSandwich, sjPlot, haven, xtable)
```

Pregunta 1

a.

La ecuación del modelo A es:

$$nvotantesy = \beta_0 + \beta_1 xpresenciamilitar * \beta_2 victima10k + \mu_i \quad (Modeloa)$$

La ecuación del modelo B es:

$$nvotantesy = \beta_0 + \beta_1 xpresenciamilitar * \beta_2 victima10k + \mu_i \quad (Modelob)$$

b.

La figura 1 Grafo de efectos fijos nos muestra la composición del modelo B.

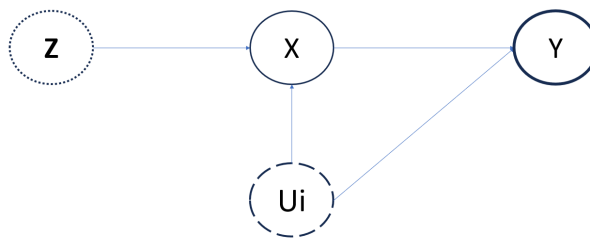


Figure 1: Grafo de efectos fijos

La composición de cada variable es:

- La Variable Z (Instrumental), es la presencia de militares
- La Variable Ui , son los valores no observados en el tiempo

- La variable X es la represión durante la dictadura
- La variable Y es el triunfo del No en 1988.

c.

Los dos supuestos centrales es que la variable instrumental: (1) Relevante = La presencia de bases militares y su distancia, está correlacionada con X que es la exposición a la represión durante dictadura. Bajo el mecanismo explicativo que estar cerca de las bases militares entrega información sobre la represión durante ese periodo, por ende, la gente es más consciente de lo que está sucediendo. (2) Exógena = La variable instrumental de presencia y distancia de las bases militares no está correlacionada con la variable independiente. El artículo sustenta que no existe correlación mencionando que la existencia de las bases son previas dictadura, por lo tanto, no fueron creadas con propósitos represivos.

Pregunta 2

a.

```
# (1) controlando por todas las covariables, salvo Pop70_pthousands
m1<- lm_robust(Share_reg70_w2 ~ LnDistMilitaryBase + share_allende70 + lnDistRegCapital + sh_rural_70 +
              weights = Pop70, data = datos)
# (2) controlando por todas las covariables
m2<- lm_robust(Share_reg70_w2 ~ LnDistMilitaryBase + share_allende70 + lnDistRegCapital + sh_rural_70 +
              weights = Pop70, data = datos)
```

b.

```
##
## =====
##               Model 1      Model 2
## -----
## (Intercept)      87.59 ***   138.52 ***
##                (11.89)      (7.70)
## LnDistMilitaryBase  -3.29      -3.61 **
##                (2.02)      (1.39)
## share_allende70   -0.26      -0.37 **
##                (0.26)      (0.11)
## lnDistRegCapital    2.38      -3.19 *
##                (2.10)      (1.39)
## sh_rural_70       -4.86     -49.71 ***
##                (11.23)     (7.31)
## IDProv            -0.37        0.38
##                (0.34)      (0.24)
## Pop70_pthousands          -30.56 ***
##                      (2.39)
## -----
## R^2                0.06        0.57
## Adj. R^2           0.04        0.57
## Num. obs.          276         276
## RMSE               4206.67     2830.52
## =====
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

El coeficiente para el logaritmo de la distancia de una base militar para el modelo 1 que no incluye una covariable en su estimación es -3,29 y no es estadísticamente significativa. Por ende, se concluye que la distancia de las bases militares no afecta los resultados del plebiscito del 88'. Sin embargo, el coeficiente para esta variable en el modelo 2 que incluye todas las covariables, si es estadísticamente significativo al 99,9% de confianza y adquiere un valor de -3,61. Sustantivamente, una duplicación en la distancia de la base militar más cercana disminuye en 3,6 puntos el registro de votantes para el plebiscito del 88'. La covariable que marco la diferencia entre ambos modelos fue Pop70_pthousands, que corresponde al total de la población dividido en 100.000. Intuitivamente, es necesario controlar por esta variable ya que la cantidad de personas registradas para votar esta estrechamente correlacionada con el total de posibles personas que pueden votar. Por ejemplo, si en un país solo votaron 100 personas, diríamos que la participación fue baja, pero si esas 100 personas son el total de la población entonces el porcentaje de participación es del 100%. De este modo, esta variable omitida estaba generando un problema de confounder, aumentando los errores estándares y subestimando el coeficiente.

c.

##Panel B-Modelo 2

```
late1 <- iv_robust(Share_reg70_w2 ~ shVictims_70 + share_allende70 + lnDistRegCapital + sh_rural_70 +
                    DMilitaryPresence + + share_allende70 + lnDistRegCapital + sh_rural_70 + IDProv,
                    weights = Pop70, data = datos)

late2 <- iv_robust(Share_reg70_w2 ~ shVictims_70 + share_allende70 + lnDistRegCapital + sh_rural_70 +
                    DMilitaryPresence + + share_allende70 + lnDistRegCapital + sh_rural_70 + Pop70_p
                    weights = Pop70, data = datos)
```

#screenreg(list(late1, late2), digits=2, include.ci=F)

```
screenreg(list(m1, m2), digits=2, include.ci=F)
```

```
##
## =====
##               Model 1      Model 2
## -----
## (Intercept)      87.59 ***   138.52 ***
##                (11.89)      (7.70)
## LnDistMilitaryBase  -3.29      -3.61 **
##                (2.02)      (1.39)
## share_allende70   -0.26      -0.37 **
##                (0.26)      (0.11)
## lnDistRegCapital    2.38      -3.19 *
##                (2.10)      (1.39)
## sh_rural_70       -4.86     -49.71 ***
##                (11.23)      (7.31)
## IDProv            -0.37        0.38
##                (0.34)      (0.24)
## Pop70_pthousands          -30.56 ***
##                        (2.39)
## -----
## R^2                0.06        0.57
## Adj. R^2           0.04        0.57
## Num. obs.          276        276
## RMSE               4206.67     2830.52
## =====
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

d.

El coeficiente de víctimas por 10.000 habitantes para el modelo 1 es 31.26 y no es estadísticamente significativo. Asimismo, este coeficiente para el modelo 2 es 14.50 y tampoco es estadísticamente significativo. Por ende, es posible concluir que el aumento en la victimización civil durante dictadura no afecta causalmente la tasa de registros de votantes para el plebiscito del 88'. Sin embargo, es difícil pensar que la tasa de victimización no afecte el registro de votantes. Por lo tanto, surge la posibilidad que existan compliers, es decir, que existe un subgrupo de la población que solo recibe el tratamiento cuando se cumplen con ciertas condiciones.

e.

Las estimaciones en el modelo A solamente buscan averiguar si existe una asociación entre la distancia de las bases militares y el registro para votar en el plebiscito del 88', pero no asume una relación causal, ya que no indica que el motivo por el cual las personas se registraron sea la distancia de las bases militares. En cambio, el modelo estimado en C estamos incluyendo una variable instrumental, en este caso, la presencia militar que es una variable dicotómica. Asumiendo el supuesto de restricción adicional que establece que la presencia de militares afecta el registro de personas para el plebiscito del 88' exclusivamente a través de una mayor exposición a la represión

Pregunta 3

a.

El argumento central del artículo desarrollado por Pepinski et al. (2023) es que controlar por efectos fijos por provincias no genera sesgos en la estimación causal entre la cercanía de los campos de concentración durante el régimen nazi y la intolerancia con grupos externos, resentimiento hacia los inmigrantes y apoyo para partidos de extrema derecha. El primer supuesto para demostrarlo es que las fronteras de los estados federales en Alemania no son un efecto o una variable postratamiento de la distancia de los campos de concentración nazi. Por lo tanto, no hay un efecto del tratamiento o la variable independiente sobre el confounder, por ende, no es un collider que sesga la estimación. El segundo argumento es que las fronteras de los estados federales en Alemania no es una variable descendiente de variable no observadas que también afectan al tratamiento (cercanía de campos de concentración nazi), así como tampoco es descendiente de variables no observadas que también afectan el resultado (intolerancia con grupos externos, resentimiento hacia inmigrantes y apoyo para partidos de extrema derecha), por ende, esta variable no genera un sesgo tipo M, que significa controlar por una variable instrumental que abre puertas traseras que estaban bloqueadas antes de añadirla al modelo.

Por lo tanto, controlar por variables post tratamiento solo generan sesgo si son descendientes de la variable independiente. Sin embargo, si no controlamos por variables no observadas que están relacionadas con características particulares de cada estado, provincia o sitio, que no son un efecto causal o descendientes del tratamiento, las estimaciones podrían estar sesgadas, ya que el término error de la variable dependiente estaría correlacionada con estas características económicas, políticas, culturales o sociales específicas de cada territorio y los resultados no serían un efecto causal directo de la variable independiente.

b1.

ww

b2.

ww