



INSTITUTO DE SOCIOLOGÍA
FACULTAD DE CIENCIAS SOCIALES

Tarea N°2

Inferencia Causal - SOL3063

Estudiante [Andreas Laffert](#)

Profesor Luis Maldonado

Ayudante Gustavo Ahumada

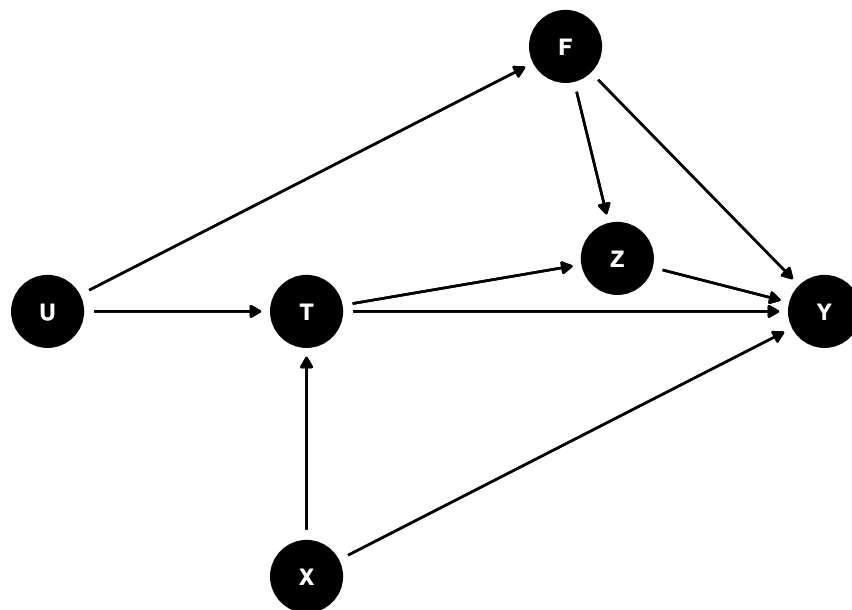
lunes 24, junio 2024

Pregunta 1

Pepinsky et al. (2024) sostienen que “The causal effect of Distance to nearest camp in Figure 3 is still only identifiable with Länder fixed effects” (p5). En base a este texto, explique esta idea.

Figura 1

Figure 1: DAG de la Figura 3 de Pepinsky et al. (2024)



Pregunta 2

Tenemos por objetivo evaluar el efecto de la intolerancia en el apoyo a partidos de extrema derecha. Para ello, utilice la distancia a campos de concentración como instrumento. Específicamente, estime los siguientes estimands:

- ITTD sin covariables y ITTD controlando por efectos fijos para Länders.
- ITT sin covariables y ITT controlando por efectos fijos para Länders.
- CACE sin covariables y CACE controlando por efectos fijos para Länders.

Reporte sus estimaciones en dos tablas de calidad: una para estimaciones sin covariables y otra con las estimaciones que controlan por FE para Länders.

Pregunta 3

En relación con sus estimaciones en el ítem 4:

- Interprete sus estimaciones: ITTD, ITT, y CACE. Además, mencione quiénes serían los cumplidores.

Table 1: Modelos sin efectos fijos para apoyo a partidos de extrema derecha en base a Pepinsky et al. (2023)

	Modelos VI		
	<i>ITT</i>	<i>ITT_D</i>	<i>CACE</i>
Distancia al campo	−0.001* (0.001)	−0.009** (0.003)	
Intolerancia			0.154 (0.079)
Länder FE	No	No	No
R ²	0.002	0.005	−0.206
Adj. R ²	0.002	0.005	−0.207
Num. obs.	2075	2075	2075
RMSE	0.153	0.685	0.168

Nota: Errores estándares robustos entre paréntesis. ** $p < 0.01$; * $p < 0.05$

Table 2: Modelos con efectos fijos para apoyo a partidos de extrema derecha en base a Pepinsky et al. (2023)

	Modelos VI		
	<i>ITT</i>	<i>ITT_D</i>	<i>CACE</i>
Distancia al campo	0.001 (0.001)	0.004 (0.004)	
Intolerancia			0.148 (0.307)
Länder FE	Sí	Sí	Sí
R ²	0.016	0.039	−0.164
Adj. R ²	0.009	0.031	−0.173
Num. obs.	2075	2075	2075
RMSE	0.153	0.676	0.166

Nota: Errores estándares robustos entre paréntesis. ** $p < 0.01$; * $p < 0.05$

Hacer tmb analisis de cumplimiento del instrumento o al menos mencionarlo:

Relevante Exogeno Restriccion de exclusion

- b) Discuta tres supuestos de sus estimaciones de CACEs, ¿son realistas? ¿se podría señalar que los CACEs estimados son creíbles.

Pregunta 4

Utilizando intolerance como variable resultado, estime los modelos 5 y 6 de Table 1 en Pepinsky et al. (2023). Específicamente:

- a) Reporte sus resultados en una tabla de calidad similar a la Table 1 del artículo bajo

replicación. Use las covariables mencionadas arriba en esta pauta¹.

Table 3: Replicación de modelos de efectos directos controlados en base a Pepinsky et al. (2023)

	Modelo 1	Modelo 2
Distancia al campo	−0.016** (0.003)	0.006 (0.005)
% Judíos (1925)	−0.897 (1.341)	5.924 (4.412)
% Desempleo (1933)	2.411** (0.589)	1.075 (0.744)
Participación partido nazi (1933)	−0.156 (0.213)	−0.717** (0.257)
Länder FE	No	Sí
Método	G-est	G-est
Num. obs.	1376.000	1376.000

Nota: Errores estándares entre paréntesis. ** $p < 0.01$; * $p < 0.05$

- b) En base a la Table 1 de Pepinsky et al. (2023), discuta dos diferencias conceptuales entre el modelo (5) y el modelo (3).
- c) Utilizando como ejemplo sus estimaciones del modelo (6), explique las nociones de efecto directo controlado promedio (ACDE) y efecto directo natural promedio (ANDE).

Pregunta 5

En nota al pie 2, Pepinsky et al. (2024) sostienen:

Sequential g-estimation, which HPT propose as a solution because they believe that Länder fixed effects are post-treatment variables, is unnecessary. It is also probably biased, because sequential g-estimation also requires different assumptions that they neither acknowledge nor defend, and only identifies a specific causal effect of T under the assumption of sequential unconfoundedness (see Acharya et al., 2016, p. 519), which would require (among other things) that there are no unobserved confounders of the causal relationship between states and contemporary tolerance. Without that assumption, about which HPT are silent, sequential g-estimation does not resolve any identification problem.

Al respecto, explique por qué la estimación-g secuencial que proponen Homola et al. (2024) sería probablemente sesgada.

¹Note que no debe incluir como variable independiente la población en 1925.

Referencias

- Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review*, 110(3), 512–529. <https://doi.org/10.1017/S0003055416000216>
- Homola, J., Pereira, M. M., & Tavits, M. (2024). Fixed Effects and Post-Treatment Bias in Legacy Studies. *American Political Science Review*, 118(1), 537–544. <https://doi.org/10.1017/S0003055423001351>
- Pepinsky, T. B., Goodman, S. W., & Ziller, C. (2023). Modeling Spatial Heterogeneity and Historical Persistence: Nazi Concentration Camps and Contemporary Intolerance. *American Political Science Review*, 118(1), 519–528. <https://doi.org/10.1017/S0003055423000072>
- Pepinsky, T. B., Goodman, S. W., & Ziller, C. (2024). Causation and History in Legacy Studies: A Reply to Homola, Pereira, and Tavits. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4705690>

Código de R

```
knitr::opts_chunk$set(echo = F,
                      warning = F,
                      error = F,
                      message = F)
if (!require("pacman")) install.packages("pacman")

pacman::p_load(tidyverse,
              rio,
              sjmisc,
              AER,
              sandwich,
              DirectEffects,
              estimatr,
              panelr,
              plm,
              clubSandwich,
              texreg,
              ggeffects)

options(scipen=999)
rm(list = ls())
db_or <- rio::import(file = "../input/data/EVS_main.csv") %>%
  as_tibble()

miles <- function(x) {
  format(round(as.numeric(x), 0), big.mark = ".")
}

decimales <- function(x) {
  format(round(as.numeric(x), 2), decimal.mark = ",")
}

custom_extract <- function(model) {
  tr <- extract(model)

  # Identificar índices a conservar (excluyendo "R$^2$", "s_idios" y "s_i
  gof_indices <- which(!(tr@gof.names %in% c("R$^2$", "s_idios", "s_id")))

  # Actualizar gof, gof.names y gof.decimal simultáneamente
```

```

tr@gof.names <- tr@gof.names[gof_indices]
tr@gof <- tr@gof[gof_indices]
tr@gof.decimal <- tr@gof.decimal[gof_indices]

return(tr)
}
# set theme

theme_set(theme_bw())

library(dagitty) # Instalar V8
library(ggdag)

coords <- list(
  x = c(Y=2, T=0, U=-1, F=1, X=0, T1=-1, T2=0, Z=1.2),
  y = c(Y=1, T=1, U=1, F=2, X=0, T1=-1, T2=-1, Z=1.2)
)

HPTbestcase <- dagify(
  Y ~ T + F + X + Z,
  F ~ U,
  Z ~ T + F,
  T ~ U + X,
  exposure = "T",
  outcome = "Y",
  coords = coords
)

tidy_dag <- tidy_dagitty(HPTbestcase)

tidy_dag$data$time <- c("t3", "t3", "t2", "t2", "t1", "t1", "t2", "t2", "

ggdag(tidy_dag) +
  geom_dag_edges() +
  geom_dag_text() +
  theme_dag()

# Seleccionar: No

db <- db_or %>%

```

```

janitor::clean_names()

# Filtrar: No

# Recodificar
sjmisc::frq(db$state)
db$state <- as.factor(db$state)

# Tratamiento casos perdidos
colSums(is.na(db)) # no NA

# Transformar y derivar: No

# ITT_Y (Z --> Y)

m1_itt <- lm_robust(formula = far_right ~ distance,
                   se_type = "HC2",
                   data = db)
m2_itt <- lm_robust(formula = far_right ~ distance,
                   fixed_effects = ~ state,
                   se_type = "HC2",
                   data = db)

# IIT_D (Z --> D)

m1_itttd <- lm_robust(formula = intolerance ~ distance,
                    se_type = "HC2",
                    data = db)
m2_itttd <- lm_robust(formula = intolerance ~ distance,
                    fixed_effects = ~ state,
                    se_type = "HC2",
                    data = db)

# CACE

m1_iv <- iv_robust(formula = far_right ~ intolerance | distance,
                  se_type = "HC2",
                  data = db,
                  diagnostics = T)
m2_iv <- iv_robust(formula = far_right ~ intolerance | distance,
                  fixed_effects = ~ state,

```



```

        se_type = "HC2",
        data = db,
        diagnostics = T)

models1 <- list(m1_itt, m1_itttd, m1_iv)
models2 <- list(m2_itt, m2_itttd, m2_iv)

ccoeff <- list(
  distance = "Distancia al campo",
  intolerance = "Intolerancia")

texreg::texreg(l = models1,
  include.ci = F,
  caption = paste("\\\\#tab:table1)", "Modelos sin efectos fij",
  stars = c(0.05, 0.01),
  custom.coef.map = ccoef,
  custom.note = "Nota: Errores estándares robustos entre par",
  custom.header = list("Modelos VI" = 1:3),
  custom.model.names = c("$ITT$", "$ITT_D$", "$CACE$"),
  leading.zero = T,
  float.pos = "h!",
  use.packages = F,
  booktabs = TRUE,
  scalebox = 1,
  digits = 3,
  custom.gof.rows = list("Länder FE" = c("No", "No", "No")))

ccoeff <- list(
  distance = "Distancia al campo",
  intolerance = "Intolerancia")

texreg::texreg(l = models2,
  include.ci = F,
  caption = paste("\\\\#tab:table1)", "Modelos con efectos fij",
  stars = c(0.05, 0.01),
  custom.coef.map = ccoef,
  custom.note = "Nota: Errores estándares robustos entre par",
  custom.header = list("Modelos VI" = 1:3),

```

```

custom.model.names = c("$ITT$", "$ITT_D$", "$CACE$"),
leading.zero = T,
float.pos = "h!",
use.packages = F,
booktabs = TRUE,
scalebox = 1,
digits = 3,
custom.gof.rows = list("Länder FE" = c("Sí", "Sí", "Sí"))

m1_direct <- sequential_g(formula = intolerance ~ distance + # A
                           prop_jewish25 + # X
                           unemployment33 + # X
                           nazishare33 | # X
                           female + # Z
                           age + # Z
                           west | # Z
                           lr + # M
                           immigrants07 + # M
                           unemployment07 + # M
                           unemp + # M
                           educ + # M
                           urban_scale, # M
                           data = db)

m2_direct <- sequential_g(formula = intolerance ~ distance + # A
                           prop_jewish25 + # X
                           unemployment33 + # X
                           nazishare33 + # X
                           state | # X (state fe as pre-treatment)
                           female + # Z
                           age + # Z
                           west | # Z
                           lr + # M
                           immigrants07 + # M
                           unemployment07 + # M
                           unemp + # M
                           educ + # M
                           urban_scale, # M
                           data = db)

```

```

# extraer para tablas

summary_m1_direct <- summary(m1_direct)
coefnames <- rownames(summary_m1_direct$coefficients)
estimates <- summary_m1_direct$coefficients[, "Estimate"]
stderrors <- summary_m1_direct$coefficients[, "Std. Err."]
pvalues <- summary_m1_direct$coefficients[, "Pr(>|t|)"]
n_obs <- as.integer(length(m1_direct$fitted.values))

m1_acde <- createTexreg(
  coef.names = coefnames,
  coef = estimates,
  se = stderrors,
  pvalues = pvalues,
  gof.names = c("Num. obs."),
  gof = as.integer(c(n_obs))
)

summary_m2_direct <- summary(m2_direct)
coefnames <- rownames(summary_m2_direct$coefficients)
estimates <- summary_m2_direct$coefficients[, "Estimate"]
stderrors <- summary_m2_direct$coefficients[, "Std. Err."]
pvalues <- summary_m2_direct$coefficients[, "Pr(>|t|)"]
n_obs <- as.integer(length(m2_direct$fitted.values))

m2_acde <- createTexreg(
  coef.names = coefnames,
  coef = estimates,
  se = stderrors,
  pvalues = pvalues,
  gof.names = c("Num. obs."),
  gof = as.integer(c(n_obs))
)

cccoef <- list(
  distance = "Distancia al campo",
  prop_jewish25 = "% Judíos (1925)",
  unemployment33 = "% Desempleo (1933)",
  nazishare33 = "Participación partido nazi (1933)"
)

```

```

)

texreg::texreg(l = list(m1_acde, m2_acde),
  caption = paste("\\#tab:table3)", "Replicación de modelos",
  stars = c(0.05, 0.01),
  custom.note = "Nota: Errores estándares entre paréntesis.",
  custom.model.names = c("Modelo 1", "Modelo 2"),
  custom.coef.map = ccoef,
  leading.zero = T,
  float.pos = "h!",
  use.packages = F,
  booktabs = TRUE,
  scalebox = 0.90,
  digits = 3,
  custom.gof.rows = list("Länder FE" = c("No", "Sí"),
    "Método" = rep("G-est", 2)))

```