

EVALUACIÓN DE IMPACTO - TALLER 1

VIAFARA MORALES, JORGE ELIECER

2025-08-20

Primer Punto

La especificación principal del autor consiste en regresar la proporción de estudiantes que hicieron trampa en el colegio c , en el grado g , del municipio m en el año t ($prop$) contra:

- a. Una dicótoma que toma el valor de uno si hubo corrupción en el municipio m en el año t ($Corrupt$).
- b. Una dicótoma que toma el valor de uno si fueron publicados reportes de auditorías en el municipio m en el año t ($Auditada$).
- c. Efectos fijos de colegio ($clavedelaescuela$).
- d. Efectos fijos de tiempo ($year$).
- e. El grado de los estudiantes ($grade$).
- f. El partido político activo ($PartidoDesf$).
- g. Una dicótoma que toma el valor de uno si el municipio fue auditado en el pasado ($AlreadyAudited$).
- h. Una dicótoma que toma el valor de uno si el municipio fue corrupto en el pasado ($CorruptPast$).
- i. El número de homicidios per cápita a nivel de municipio-año ($HOMI_CAP_MUN$).
- j. El logaritmo del total de impuestos recolectados en el municipio m en el año t ($total$).
- k. Una dicótoma que toma el valor de uno si el partido político del municipio m en el año t está alineado con el nacional ($MismoPartidoG$).

Escriban matemáticamente la regresión a estimar y expliquen claramente a qué corresponde cada término. Asegúrense de escribir claramente a qué nivel de agregación está cada una de las variables incluidas en la regresión. Así mismo, indiquen cuál es el parámetro de interés el cual responde la pregunta de investigación del autor. De ser preferible, pueden resumir las variables de control – y solo las variables de control – en un único vector, sin embargo, deben explicar qué contiene dicho vector y cuál es su nivel de agregación.

Solución Primer Punto

De acuerdo con el enunciado la especificación principal del autor puede escribirse de la siguiente manera:

$$\begin{aligned} prop_{cgmt} = & \alpha + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + Clavedelaescuela_c \\ & + year_t + grade_{gmt} + X_{mt} + u_{cgmt} \end{aligned} \quad (1)$$

Nota: la nomenclatura utilizada en la ecuación (1) sigue la convención de subíndices para indicar el nivel de agregación de las variables.

A continuación, se definen los términos de la ecuación (1):

- $prop_{cgmt}$: Esta es la variable dependiente que estima el resultado potencial de aquellos estudiantes que hicieron trampa en el colegio (c), en el grado (g) del municipio (m) en el año (t). El nivel de agregación de la variable es colegio, grado, municipio y año.
- $Corrupt_{mt}$: Esta es una variable dicotómica que toma el valor de uno si hubo corrupción en el municipio (m) en el año (t). El nivel de agregación es municipio y año.
- $Auditada_{mt}$: Esta es una variable dicotómica que toma el valor de uno si fueron publicados reportes de auditorías en el municipio (m) en el año (t). El nivel de agregación es municipio y año.
- $Clavedelaescuela_c$: Esta es una variable de efectos fijos de colegio, que captura las características constantes de cada colegio (c) a lo largo del tiempo. El nivel de agregación es colegio.
- $year_t$: Esta es una variable de efectos fijos de tiempo, que captura las características constantes a lo largo del tiempo (t). El nivel de agregación es año.
- $grade_{gmt}$: Esta es una variable que representa el grado de los estudiantes (g) en el municipio (m) en el año (t). El nivel de agregación es grado, municipio y año.
- X_{mt} : Este es un vector de variables de control con nivel de agregación municipio y año, que incluye:
 - $PartidoDesf_{mt}$: El partido político activo en el municipio (m) en el año (t).
 - $AlreadyAudited_{mt}$: Una variable dicotómica que toma el valor de uno si el municipio fue auditado en el pasado.
 - $CorruptPast_{mt}$: Una variable dicotómica que toma el valor de uno si el municipio fue corrupto en el pasado.
 - $HOMI_CAP_MUN_{mt}$: El número de homicidios per cápita a nivel de municipio-año.
 - $\log(total)_{mt}$: El logaritmo del total de impuestos recolectados en el municipio (m) en el año (t).
 - $MismoPartidoG_{mt}$: Una variable dicotómica que toma el valor de uno si el partido político del municipio (m) en el año (t) está alineado con el nacional.
 - u_{cgmt} : Este es el término de error que captura las variables no observadas que afectan la variable dependiente (prop) a nivel de colegio (c), grado (g), municipio(m) y año(t).

Manual avanzado de R markdown: <https://rpubs.com/ricardo/14631>

Segundo Punto

Corra las siguientes estimaciones y en una tabla presente los resultados de las siguientes regresiones

- Una regresión simple de la variable dependiente contra la independiente principal.
- Una regresión simple de la variable dependiente contra la independiente principal + Controles.
- Una regresión simple de la variable dependiente contra la independiente principal + Efectos fijos.
- Una regresión simple de la variable dependiente contra la independiente principal + Controles + Efectos fijos.

La tabla debe tener el formato de presentación tipo artículo y estar completamente en español. Asegúrense de que esta tabla presente únicamente el coeficiente asociado a la variable de interés – i.e., no presente los coeficientes asociados a los controles, el intercepto o los efectos fijos. Debe ser claro, sin embargo, qué es incluido en cada columna. Para esto, pueden seguir el siguiente formato de presentación.

Generalidades de R

```

#Limpiar la consola
#cat("\f")

#Limpiar el Global Environment
#rm(list = ls())

#Incluir las librerías
if(!require(pacman)) install.packages("pacman") ; require(pacman)

```

```
## Loading required package: pacman
```

```

p_load(haven,    #Leer archivos .dta
       dplyr,    #Manipular datos
       stargazer, #Visualizar tablas de regresión
       fixest,   #Calcular los efectos fijos
       plm,      #Datos de panel
       knitr     #Visualizar tablas adicionales
)

```

Transformación de datos en R

```

# Definir URL del repositorio
github_url <- "https://raw.githubusercontent.com/GeorgeWton1986/Eva_Impacto_Meca/main"

# Cargar datos
corruption_data <- read_dta(paste0(github_url, "/Data/corruption_SV_github.dta"))

# Ver estructura de los datos
#glimpse(corruption_data)
#head(corruption_data)

#Selección de columnas
corruption_1 <- corruption_data %>%
  select(prop, Corrupt, Auditada, clavedelaescuela, year,
         GradoSecundaria, PartidoDesf, AlreadyAudited,
         CorruptPast, HOMI_CAP_MUN, total, MismoPartidoG)

#Transformar la variable total a logaritmo natural
corruption_1$log_total <- log(corruption_1$total + 1)

#Limpiar la base de datos corruption_1 y obtener la base de datos corruption_clean
corruption_clean <- corruption_1 %>%
  filter(!is.na(prop), !is.na(Corrupt), !is.na(Auditada))

# Verificar datos limpios
cat("Observaciones después de limpiar NA's:", nrow(corruption_clean))

```

```
## Observaciones después de limpiar NA's: 102133
```

```
#Verificacion de variables
summary(corruption_clean)
```

```
##      prop      Corrupt      Auditada      clavedelaescuela
## Min.   :0.00000 Min.   :0.0000 Min.   :0.0000 Length:102133
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 Class :character
## Median :0.00000 Median :0.0000 Median :0.0000 Mode  :character
## Mean   :0.04107 Mean   :0.1535 Mean   :0.3245
## 3rd Qu.:0.01843 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max.   :1.00000 Max.   :1.0000 Max.   :1.0000
##      year      GradoSecundaria PartidoDesf      AlreadyAudited
## Min.   :2006 Min.   :1.000 Length:102133 Min.   :0.0000
## 1st Qu.:2009 1st Qu.:1.000 Class :character 1st Qu.:0.0000
## Median :2011 Median :2.000 Mode  :character Median :1.0000
## Mean   :2010 Mean   :2.145 Mean   :0.7077
## 3rd Qu.:2012 3rd Qu.:3.000 3rd Qu.:1.0000
## Max.   :2013 Max.   :3.000 Max.   :1.0000
## CorruptPast      HOMI_CAP_MUN      total      MismoPartidoG
## Min.   :0.0000 Min.   :0.000e+00 Min.   :0.000e+00 Min.   :0.0000
## 1st Qu.:0.0000 1st Qu.:4.746e-05 1st Qu.:1.460e+08 1st Qu.:0.0000
## Median :0.0000 Median :1.111e-04 Median :3.758e+08 Median :0.0000
## Mean   :0.4692 Mean   :2.163e-04 Mean   :1.005e+09 Mean   :0.3178
## 3rd Qu.:1.0000 3rd Qu.:2.235e-04 3rd Qu.:1.488e+09 3rd Qu.:1.0000
## Max.   :1.0000 Max.   :4.556e-03 Max.   :7.762e+09 Max.   :1.0000
## log_total
## Min.   : 0.00
## 1st Qu.:18.80
## Median :19.74
## Mean   :19.13
## 3rd Qu.:21.12
## Max.   :22.77
```

```
head(corruption_clean)
```

```
## # A tibble: 6 x 13
##      prop Corrupt Auditada clavedelaescuela year GradoSecundaria PartidoDesf
##      <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr>
## 1 0.00893 0 0 01DES00010 2006 3 PAN
## 2 0 0 0 01PES0037H 2006 3 PAN
## 3 0 0 0 01DTV0095Z 2006 3 PAN
## 4 0 0 0 01PES0004Q 2006 3 PAN
## 5 0.00901 0 0 01DST0035X 2006 3 PAN
## 6 0.0244 0 0 01PES0046P 2006 3 PAN
## # i 6 more variables: AlreadyAudited <dbl>, CorruptPast <dbl>,
## # HOMI_CAP_MUN <dbl>, total <dbl>, MismoPartidoG <dbl>, log_total <dbl>
```

Solución Punto a)

Una regresión simple de la variable dependiente contra la independiente principal. En este sentido, el modelo_a, se especifica de la siguiente forma:

$$prop_{cgmt} = \alpha + \beta_1 Corrupt_{mt} + u_{cgmt} \quad (2)$$

```
#Regresion del modelo simple.
```

```
modelo_a <- lm(prop ~ Corrupt,
               data = corruption_clean)
```

Solución Punto b)

Una regresión simple de la variable dependiente contra la independiente principal + Controles. Por lo tanto, el modelo_b se especifica de la siguiente forma:

$$\begin{aligned} prop_{cgmt} = & \alpha + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + grade_{gmt} \\ & + PartidoDesf_{mt} + AlreadyAudited_{mt} + CorruptPast_{mt} \\ & + HOMICAPMUN_{mt} + MismoPartidoG_{mt} + \log(total)_{mt} + u_{cgmt} \end{aligned} \quad (3)$$

```
#Regresion del modelo simple de la variable dependiente contra  
#la independiente principal + Controles.
```

```
modelo_b <- lm(prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf  
               + AlreadyAudited + CorruptPast + HOMI_CAP_MUN + log_total + MismoPartidoG,  
               data = corruption_clean)
```

Solución Punto c)

Una regresión simple de la variable dependiente contra la independiente principal + Efectos fijos.

$$\begin{aligned} prop_{cgmt} = & \alpha + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + Clavedelaescuela_c \\ & + year_t + u_{cgmt} \end{aligned} \quad (4)$$

```
#Convertir corruption_clean a un data frame de panel
```

```
corruption_panel <- pdata.frame(corruption_clean, index = c("clavedelaescuela", "year"))
```

```
## Warning in pdata.frame(corruption_clean, index = c("clavedelaescuela", "year")): duplicate couples (
## to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
#Una regresión simple de la variable dependiente contra  
#la independiente principal + Efectos fijos.
```

```
modelo_c <- plm(prop ~ Corrupt + Auditada, data = corruption_panel,  
               model = "within",  
               effect = "twoways")
```

Solución Punto d)

Una regresión simple de la variable dependiente contra la independiente principal + Controles + Efectos fijos.

```
#Una regresión simple de la variable dependiente
#contra la independiente principal + Controles + Efectos fijos.

modelo_d <- plm(prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf
               + AlreadyAudited + CorruptPast + HOMI_CAP_MUN + log_total + MismoPartidoG,
               data = corruption_panel,
               model = "within",
               effect = "twoways")
```

Tabla Resumen de Resultados

```
# Opción 2: Con ajustes LaTeX adicionales para forzar el ajuste
stargazer(modelo_a, modelo_b, modelo_c, modelo_d,
           type = "latex",
           title = "Efecto de la Corrupción sobre la Proporción de Trampa Estudiantil",
           column.labels = c("(1) Simple", "(2) Controles", "(3) EF", "(4) Completo"),
           omit = c("Auditada", "GradoSecundaria", "PartidoDesf",
                   "AlreadyAudited", "CorruptPast", "HOMI_CAP_MUN",
                   "log_total", "MismoPartidoG", "Constant"),
           covariate.labels = c("Corrupción"),
           dep.var.labels = "Proporción de Trampa",
           add.lines = list(
             c("Controles", "No", "Sí", "No", "Sí"),
             c("EF Colegio", "No", "No", "Sí", "Sí"),
             c("EF Año", "No", "No", "Sí", "Sí")
           ),
           notes = c("Errores estándar en paréntesis.",
                    "*** p<0.01, ** p<0.05, * p<0.1"),
           notes.append = FALSE,
           font.size = "scriptsize", # Fuente más pequeña
           column.sep.width = "1pt", # Separación mínima
           no.space = TRUE,
           header = FALSE,
           table.placement = "htbp", # Cambiar de "H" a "htbp"
           float = TRUE,
           float.env = "table")
```

Tercer Punto

Interpreten los resultados encontrados a partir de la primera regresión (estimación punto 2, inciso b). ¿Parece haber una relación entre la corrupción política y la trampa de los estudiantes? ¿Podemos interpretar esta relación como causal? ¿Cuál sería el supuesto de identificación? Argumenten.

Solución Tercer Punto

¿Parece haber una relación entre la corrupción política y la trampa de los estudiantes? ¿Podemos interpretar esta relación como causal? ¿Cuál sería el supuesto de identificación? Argumenten

Respuesta:

Table 1: Efecto de la Corrupción sobre la Proporción de Trampa Estudiantil

	<i>Dependent variable:</i> Proporción de Trampa			
	<i>OLS</i>			<i>panel linear</i>
	(1) Simple (1)	(2) Controles (2)	(3) EF (3)	(4) Completo (4)
Corrupción	0.001 (0.001)	0.001 (0.001)	0.006*** (0.001)	0.003** (0.002)
Controles	No	Sí	No	Sí
EF Colegio	No	No	Sí	Sí
EF Año	No	No	Sí	Sí
Observations	102,133	102,133	102,133	102,133
R ²	0.00001	0.031	0.0002	0.013
Adjusted R ²	0.00000	0.030	-0.250	-0.235
Residual Std. Error	0.126 (df = 102131)	0.124 (df = 102043)		
F Statistic	1.503 (df = 1; 102131)	36.722*** (df = 89; 102043)	8.191*** (df = 2; 81663)	12.470*** (df = 89; 81576)

Note:

Errores estándar en paréntesis.
*** p<0.01, ** p<0.05, * p<0.1

En primer lugar, es necesario definir los grupos de tratamiento y control para poder interpretar los resultados de la regresión. En este caso, el grupo de tratamiento está compuesto por los colegios ubicados en municipios que presentaron corrupción en el año t , mientras que el grupo de control está conformado por aquellos colegios situados en municipios que no presentaron corrupción en ese mismo año t .

Segundo lugar, el término supuesto de identificación, se refiere al resultado promedio de los colegios tratados si no hubieran estado expuestos a la corrupción (Contrafactual), sea igual al resultado promedio de los colegios de control. Es decir, los colegios tratados y los colegios de control deben ser comparables en la ausencia de corrupción. Este término es clave, porque se supone que los colegios tratados y control poseen las mismas características observables y no observables.

Tercer lugar, en caso de no cumplirse el supuesto de identificación, se denomina un sesgo de selección, donde existen diferencias en la características de los colegios tratados y control. Por lo tanto, los coeficientes estimados de la regresión son sesgados e inconsistentes, lo cual, puede llevar a interpretación equivocada en términos de sobre estimación o subestimar el efecto de la corrupción política sobre la trampa de los estudiantes.

De acuerdo con los resultados de los modelos econométricos, se puede mencionar que,

- **Modelo_a (Simple):** El modelo no cuenta con controles, ni con efectos fijos de colegio, ni de año. Por lo tanto, la variable corrupción (Corrupt) tiene un coeficiente de 0.001 el cual es no significativo. Además, el modelo_a está especificado para realizar una comparación entre municipios con corrupción en el tiempo t y aquellos municipios que no la tuvieron en el tiempo t . En este sentido, se puede interpretar que no existe una relación entre la corrupción política y la trampa de los estudiantes. Con respecto al supuesto de identificación, este no se cumple, los municipios tienen diferentes características en términos de campañas para la prevención de la corrupción, además, los colegios pueden tener infraestructura insuficiente, docentes con poca experiencia y formación.
- **Modelo_b (Controles):** El modelo incluye controles, sin embargo, no incluye efectos fijos de colegio, ni de año. Por lo tanto, la variable corrupción (Corrupt) tiene un coeficiente de 0.001 el cual es no significativo, este coeficiente no mejora a pesar de la inclusión de los controles. En resumen, las variables observables (Controles) son insuficientes para capturar el efecto de la corrupción política en los estudiantes. Ahora, en términos del supuesto de identificación, este no se cumple, porque continúan existiendo diferencias no observables entre los grupos, tales como cultura institucional del colegio, calidad de la gestión directiva, y características socioeconómicas del entorno escolar que no se capturan en los controles.
- **Modelo_c (Efectos Fijos):** El modelo incluye efectos fijos de colegio y de año. Sin embargo, no incluye controles. Este cambio en la especificación del modelo, permitió que el coeficiente tuviera

un valor de 0.006 y obtuviera una significancia estadística del 1%. El modelo_c comparó el mismo colegio en diferentes años cuando estuvo expuesto a la corrupción y cuando no lo estuvo. De acuerdo con lo anterior, implementar esta especificación facilitó observar que la heterogeneidad no observada entre colegios confundía la relación en los modelos anteriores, ocultando el verdadero efecto causal, e.i. elimina variables no observables como infraestructura, localización, sistemas de evaluación, entre otros. Ahora bien, el supuesto de identificación es más real porque los efectos fijos eliminan heterogeneidad no observada invariante en el tiempo, es decir, el efecto de la corrupción política sobre la trampa de los estudiantes es causal.

- **Modelo_d (Controles + Efectos Fijos):** El modelo incluye controles, efectos fijos de colegio y de año. El coeficiente de la variable corrupción (Corrupt) es 0.003 y es significativo al 5%. Este modelo se especificó de manera completa combinando los efectos fijos y los controles observables. En concordancia, el efecto fijo permite observar el mismo colegio en diferentes años y los controles ayudan a capturar efecto geográfico y político del municipio en el tiempo t. Por lo tanto, la inclusión de los controles permite explicar la variación en el coeficiente con respecto del modelo_c. Por último, el supuesto de identificación es más real, porque los colegios tratados y control son comparables en la ausencia de corrupción, además, se tienen en cuenta otras características de los municipios en tiempo.

Cuarto Punto

Escojan la especificación que – a su criterio – parece ser la más adecuada para aproximarse a la respuesta de la pregunta de investigación del autor y respondan:

a) ¿Cuál es el supuesto de identificación para que esta especificación les permita hacer interpretaciones causales?

Solución Cuarto Punto

¿Cuál es el supuesto de identificación para que esta especificación les permita hacer interpretaciones causales?

Respuesta:

El modelo_d presenta la especificación más adecuada al combinar efectos fijos y controles observables. En este sentido, este modelo permite tener un supuesto de identificación más real, debido a que se compara el mismo colegio en diferentes años bajo una condición exógena de corrupción y si no hubiera corrupción. Además, se están teniendo en cuenta las características políticas, financieras y socioeconómicas del municipio en el tiempo t. De esta manera, la heterogeneidad no observada invariante en el tiempo se elimina por causa de los efectos fijos y los controles capturan las variables observables que varían en el tiempo t. Por lo cual, hace más real la interpretación causal del efecto.

Quinto Punto

Dados los resultados de la especificación que escogieron en el anterior punto: interpreten el coeficiente estimado.

Solución Quinto Punto

¿Cómo se puede interpretar el resultado del modelo_d?

Respuesta:

La interpretación estadística del coeficiente de la variable corrupción (Corrupt) en el modelo_d indica que la corrupción municipal incrementa en 0.3 punto porcentuales la proporción de estudiantes que hicieron trampa en el colegio. En términos de p-valor, la variable corrupción es significativa al 5%, por lo tanto, se rechaza la

hipótesis nula (H_0) porque existe al menos 5% de probabilidad de observar la relación entre la corrupción y la trampa en el colegio.