

EVALUACIÓN DE IMPACTO - TALLER 4

VIÁFARA MORALES, JORGE ELIECER

2025-09-27

Introducción

En su artículo “Soap Operas and Fertility: Evidence from Brazil” , La Ferrara et al. (2012) estudian la influencia de la televisión, y específicamente de las telenovelas, sobre las preferencias de las mujeres respecto a sus preferencias sobre su vida reproductiva. En particular, aprovechan la llegada de una cadena de televisión (Globo) en ciertas áreas para estimar la probabilidad de dar a luz. Adicionalmente, evalúan la existencia de efectos heterogéneos en variables como la edad, nivel educativo y riqueza de las mujeres. Para responder sus preguntas, los autores usan la metodología de Diferencias en Diferencias con regresión y efectos fijos.

En este taller, ustedes realizarán algunas de las estimaciones hechas por los autores e interpretarán los resultados. Para esto, ustedes deben usar una submuestra aleatoria de la base de datos SoapOperas.dta, la cual fue usada por los autores. Así, justo después de abrir la base de datos (la base completa), cada grupo debe eliminar aleatoriamente el 5% de las observaciones y usar la base restante. La semilla que deben usar para que sus resultados sean replicables es su código de estudiante. Asegúrese de restringir la base de datos hasta el percentil 95 del tamaño del área (geoarea80).

Como su solución a este conjunto de problemas, proporcione un documento PDF con sus respuestas y el do file (u otro programa) que usó para resolverlo. Se recomienda tener respuestas tan breves como sea posible. Ciertas preguntas, sobre todo aquellas que preguntan por su opinión, pueden no tener una única respuesta correcta. Así, lo importante es que argumenten de manera coherente.

Primer Punto

1. Con base en el artículo descrito resuelva:

Solución Primer Punto

- a) Escriban la especificación principal de los autores (en niveles) y describan cada una de sus variables explicativas.

Respuesta: Los autores especifican el modelo de probabilidad lineal, de la siguiente manera:

$$y_{ijt} = X_{ijt}\beta + \gamma G_{jt} + \mu_j + \lambda_t + \epsilon_{ijt} \quad (1)$$

Donde: - y_{ijt} Es igual a 1 si la mujer i vive en el área j y da a luz a un niño en año t . - X_{ijt} Son los controles que varían en el tiempo del individuo y a nivel de AMC. - G_{jt} Es una dicotoma que toma el valor de 1, si el área j recibe señal de Rede Globo al menos un año antes del año t . - μ_j Efectos fijos del área - λ_t Efectos fijos de tiempo - ϵ_{ijt} es el término de error.

b) ¿Cuál es el supuesto de identificación detrás de este modelo?

Respuesta: El modelo de diferencias en diferencias (DiD) reconoce el supuesto de identificación mediante las tendencias paralelas. Esto significa, el cambio en la variable resultado para el grupo de control es igual al cambio en la variable resultado para el grupo de tratamiento, sino hubiera sido tratado. Por lo tanto, el cambio en la probabilidad de fertilidad esperada en el grupo de control (antes y después de que Globo llegue a las áreas tratadas) sería igual al cambio en la probabilidad de fertilidad esperada en el grupo tratado si este no hubiera recibido el tratamiento. $[E[y_{ijt}|G_{jt} = 0, t = 1] - E[y_{ijt}|G_{jt} = 0, t = 0]] - [[E[y_{ijt}|G_{jt} = 1, t = 1] - E[y_{ijt}|G_{jt} = 1, t = 0]]$

Generalidades de R

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

```

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

#Incluir las librerias
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
       ivreg,      #Regresiones por variables instrumentales
       broom,      #Extraer resultados de regresiones
       AER,        #Regresiones por variables instrumentales
       sandwich,   #Errores estándar robustos
       lmtest,     #Pruebas estadísticas
       kableExtra  #Tablas avanzadas
       )

```

Transformación de datos en R

```

# Definir URL del repositorio
github_url <- "https://raw.githubusercontent.com/GeorgeWton1986/Eva_impacto_T4/main"
# Cargar datos
BD_0 <- read_dta(paste0(github_url, "/Data/SoapOperas2.dta"))

# Ver estructura de los datos
#glimpse(BD_0)
#head(BD_0)

# Verificar los nombres de la columnas
colnames(BD_0)

```

```

## [1] "rural"      "weight"     "catholic"   "yrs_edu"
## [5] "married"    "yrseu_head" "amc_code"   "age"
## [9] "year"      "B"          "stock"      "globocoverage1"
## [13] "geoarea80" "ipc_renta"  "wealth_noTV" "Doctors"
## [17] "agesq"     "stocksq"    "cov1wealth" "cov1eduhd"
## [21] "cov1edu"   "treatment"  "selecc"

```

- c) Estime (1) el modelo sin controles y sin efectos fijos de área (2) sin controles, pero con efectos fijos de área (amc_code) (3) con controles y con efectos fijos de área.

Nota: para todos los modelos incluya efectos fijos de tiempo, ajuste por sobremuestreo (ponderando las observaciones por la variable weight) y clusterice errores por área. Luego presente y analice sus resultados (usando cluster () como opción de la regresión).

```

# Incluir la semilla con mi código de estudiante
set.seed(202415176)

# Eliminar aleatoriamente el 5% de las observaciones
BD_1 <- BD_0 %>% sample_frac(0.95)

# Restringir el tamaño del área hasta el percentil 95
percentil_95 <- quantile(BD_1$geoarea80, 0.95, na.rm = TRUE)
BD_1 <- BD_1 %>% filter(geoarea80 <= percentil_95)

```

Respuesta: A continuación, presentamos el desarrollo de los modelos.

```

# Modelo 1: Sin controles y sin efectos fijos de área

mod_1 <- feols(B ~ globocoverage1 | year,
               data = BD_1,
               weights = BD_1$weight,
               cluster = "amc_code")

# Modelo 2: Sin controles, pero con efectos fijos de área (amc_code)

mod_2 <- feols(B ~ globocoverage1 | year + amc_code,
               data = BD_1,
               weights = BD_1$weight,
               cluster = "amc_code")

# Modelo 3: Con controles y con efectos fijos de área

mod_3 <- feols(B ~ globocoverage1 + rural + catholic + yrs_edu + married +
               yrsedu_head + age + stock + ipc_renta + wealth_noTV + Doctors +
               agesq + stocksq | year + amc_code,
               data = BD_1,
               weights = BD_1$weight,
               cluster = "amc_code")

# Resultado "Efecto de la Cobertura de Globo en la Probabilidad de nacimientos"
etable(mod_1, mod_2, mod_3,
        headers = c("Modelo 1", "Modelo 2", "Modelo 3"),
        keep = "globocoverage1",
        digits = 4,
        se.below = TRUE)

```

```

##               mod_1      mod_2      mod_3
##           Modelo 1  Modelo 2  Modelo 3
## Dependent Var.:      B          B          B
##
## globocoverage1 -0.0168*** -0.0051** -0.0060**
##                (0.0031)  (0.0019)  (0.0019)
## Fixed-Effects:  -----  -----  -----
## year           Yes       Yes       Yes
## amc_code        No       Yes       Yes
## -----  -----  -----

```

```
## S.E.: Clustered by: amc_.. by: amc.. by: amc..
## Observations      671,302    671,302    671,153
## R2                0.00222    0.01095    0.04938
## Within R2         0.00060    1.93e-5    0.03888
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

Segundo Punto

2. Mencione dos de los retos de identificación mencionados por los autores sobre esta estimación y la forma como se puede corregir con una técnica econométrica. (**Nota:** Puede encontrarlos en la introducción del artículo, página 3).

Solución Segundo Punto

Respuesta: i) El primer reto de identificación es la posibilidad que la presencia de Globo puede estar correlacionada con determinantes no observadas de la fertilidad. La entrega de globo puede esta correlacionada con las tendencias de fertilidad preexistentes. Esto fue corregido mediante la regresión de la fertilidad sobre un conjunto de variables dummies que van desde 9 años antes y 9 años después de la entrada de Globo en las áreas estudiadas.

- ii) El segundo reto de identificación es la posibilidad de presentar una endogeniedad en el momento en que Globo entro a diferentes áreas. Entendiendo que podría ser una estrategia comercial sobre áreas ricas que pueden generar ingresos por publicidad. La corrección econométrica fue controlar por medio de la educación, riqueza y por el índice de potencial de consumo, este último es usado por Globo para evaluar los mercados nuevos. Adicionalmente, los autores probaron los resultados con pruebas de robuztes.

Tercer Punto

3. Estime el modelo (3) del literal c, restringiendo los datos a aquellas mujeres con edad entre i) 15-24 ii) 25-34 iii) 35-44 presente sus resultados en una misma tabla con cada modelo en una columna distinta (3 modelos).

Solución tercer Punto

```
# Filtrar los datos por rangos de edad

BD_15_24 <- BD_1 %>% filter(age >= 15 & age <= 24)
BD_25_34 <- BD_1 %>% filter(age >= 25 & age <= 34)
BD_35_44 <- BD_1 %>% filter(age >= 35 & age <= 44)

# Modelos con restricción de edad

mod_3_1 <- feols(B ~ globocoverage1 + rural + catholic + yrs_edu + married +
  yrsedu_head + age + stock + ipc_renta + wealth_noTV + Doctors +
  agesq + stocksq | year + amc_code,
```

```

data = BD_15_24,
weights = BD_15_24$weight,
cluster = "amc_code")

mod_3_2 <- feols(B ~ globocoverage1 + rural + catholic + yrs_edu + married +
  yrsedu_head + age + stock + ipc_renta + wealth_noTV + Doctors +
  agesq + stocksq | year + amc_code,
data = BD_25_34,
weights = BD_25_34$weight,
cluster = "amc_code")

mod_3_3 <- feols(B ~ globocoverage1 + rural + catholic + yrs_edu + married +
  yrsedu_head + age + stock + ipc_renta + wealth_noTV + Doctors +
  agesq + stocksq | year + amc_code,
data = BD_35_44,
weights = BD_35_44$weight,
cluster = "amc_code")

# Resultado "Efecto de la Cobertura de Globo en la Probabilidad de nacimientos por Rangos de Edad"

etable(mod_3_1, mod_3_2, mod_3_3,
  headers = c("Modelo 15-24", "Modelo 25-34", "Modelo 35-44"),
  keep = "globocoverage1",
  digits = 4,
  se.below = TRUE)

```

##	mod_3_1	mod_3_2	mod_3_3
##	Modelo 15-24	Modelo 25-34	Modelo 35-44
## Dependent Var.:	B	B	B
##			
## globocoverage1	-0.0039	-0.0112**	-0.0053
##	(0.0024)	(0.0041)	(0.0028)
## Fixed-Effects:	-----	-----	-----
## year	Yes	Yes	Yes
## amc_code	Yes	Yes	Yes
##	-----	-----	-----
## S.E.: Clustered	by: am..	by: amc..	by: am..
## Observations	265,927	208,816	143,235
## R2	0.07243	0.03732	0.05075
## Within R2	0.06034	0.01740	0.01654
## ---			
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Respuesta: Para el grupo de mujeres entre 15 y 34 años, la influencia por causa del ingreso de Globo no es estadísticamente significativa. El grupo de mujeres entre 25 y 34 años, la cobertura de Globo es estadísticamente significativa al 1% y estima que la probabilidad de nacimientos se reduce en 1.12 puntos porcentuales. Por último, las mujeres con un rango de edad entre 35 y 44 años, la presencia de Globo es estadísticamente significativa al 10% y calcula una disminución en 0.53 puntos porcentuales la probabilidad de nacimientos.

En conclusión: Las mujeres entre los 25 y 34 años son influenciadas por la telenovelas de Globo sobre la planificación familiar. Las mujeres más jóvenes no tienen influencia alguna por parte de las telenovelas, dado que aún no les interesa iniciar su vida reproductiva.

Cuarto Punto

4. Escoja una de las siguientes variables y estime un modelo con efectos heterogéneos:

- i) años de educación de cabeza de hogar (yrsedu_head)
- ii) años de educación de la mujer (yrs_edu)
- iii) riqueza del hogar (wealth_noTV).

Solución cuarto Punto

```
# Modelo con efectos heterogéneos con años de educación de la mujer (yrs_edu)

mod_het <- feols(B ~ globocoverage1 + globocoverage1*yrs_edu + rural + catholic + yrs_edu +
                 married + yrsedu_head + age + stock + ipc_renta + wealth_noTV + Doctors +
                 agesq + stocksq | year + amc_code,
                 data = BD_1,
                 weights = BD_1$weight,
                 cluster = "amc_code")

# Resultado "Efecto Globo en la Probabilidad de nacimientos con Efectos Heterogéneos
# por Años de Educación de la mujer"

etable(mod_het,
        keep = c("globocoverage1", "globocoverage1:yrs_edu", "yrs_edu"),
        digits = 4,
        se.below = TRUE)
```

```
##                               mod_het
## Dependent Var.:                B
##
## globocoverage1                -0.0142***
##                               (0.0026)
## yrs_edu                       -0.0050***
##                               (0.0002)
## globocoverage1 x yrs_edu      0.0016***
##                               (0.0002)
## Fixed-Effects:                -----
## year                          Yes
## amc_code                       Yes
## -----
## S.E.: Clustered               by: amc_..
## Observations                  671,153
## R2                            0.04948
## Within R2                     0.03898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Interprete los resultados del efecto heterogéneo estimado en el modelo y concluya sobre estos. (Nota: un efecto heterogéneo se estima mediante agregar al modelo estándar una interacción entre el tratamiento y la variable que se quiere estudiar, también recuerde que las variables de la interacción deben estar presentes de manera individual en el modelo)

Respuesta: En general las variables de cobertura de Globo, los años de educación de la mujer y la interacción son significativos en el estudio de efecto. Ahora, me centraré en explicar los resultados por cada uno de los coeficientes:

globocoverage1: En mujeres sin educación (0 años), la cobertura de Globo reduce la probabilidad de nacimientos en 1.42 puntos porcentuales.

yrs_edu: En área sin cobertura de Globo, cada año adicional de educación reduce la probabilidad de dar a luz en 0.05 puntos porcentuales.

globocoverage1:yrs_edu: La interacción presenta el cambio del efecto de la cobertura de Globo por cada año adicional de educación de la mujer. Es decir, el efecto de Globo reduce en 0.16 puntos porcentuales con un aumento de un año adicional en la educación de la mujer. Por lo tanto, las mujeres más educadas son menos influenciadas por las telenovelas.

En conclusión: Aumentar los años de educación en mujeres se relaciona con un efecto menor a causa de la influencia de las telenovelas. Porque, las mujeres más educadas tienen la posibilidad de informarse mejor.