

Taller 2 - Evaluación de impacto

Camila Caraballo, Laura Rivera

2025-09-02

Taller 2- Variables Instrumentales

```
# Cargar librerías y paquetes

#install.packages("haven")
#install.packages("plm")
#install.packages("skimmer")
#install.packages("dplyr")
#install.packages("ggplot2")
#install.packages("modelsummary")
library(knitr)
library(haven)
library(plm)
library(scales)
library(skimmer)
library(dplyr)
library(ggplot2)
library(AER)
library(fixest)
library(lmtest)
library(sandwich)
library(tidyverse)
library(stargazer)

# Leer archivo .dta
data <- read_dta("oil_windfall_sin10perc.dta")
set.seed(201613424)
```

1. Ecuaciones

Especificación en niveles (OLS, Ecuación 1)

$$W_m = \delta_1 + \theta_1 R_m + X_m \rho_1 + u_{m1}$$

En este modelo, W_m representa el conjunto de resultados económicos a nivel de AMC que incluyen el gasto reportado en distintas partidas de presupuesto a nivel municipal, provisión de bienes y servicios públicos, transferencias, ingresos de los hogares y tasas de pobreza. A su vez, R_m corresponde a una medida de los ingresos municipales y es la variable explicativa principal. Finalmente, X_m reúne distintas características geográficas de cada municipio, incluyendo su latitud y longitud, si está ubicado en la costa, la distancia a las capitales federal y estatal, un indicador de si es capital estatal, así como efectos fijos por estado.

Especificación VI (OLS, Ecuación 3 y 4)

$$Y_{m,1970} = \delta_3 + \eta_0 Q_{m,2000} + X'_m \rho_3 + w_{m3}$$

Donde $Y_{m,1970}$ es el resultado socioeconómico en AMC para el año 1970 y se usa como “placebo” para probar que el instrumento no estaba correlacionado con condiciones socioeconómicas históricas, $Q_{m,2000}$ hace referencia a la producción petrolera en el año 2000 en el municipio m , el cual funciona como el instrumento para regalías y X'_m es el vector de controles geográficos y estructurales como latitud, longitud, si es costa, distancias a capitales, si es capital estatal junto con efectos fijos estatales.

$$Y_{m,2002} = \delta_4 + \eta_4 Q_{m,2002} + X'_m \rho_4 + w_{m4}$$

Donde $Y_{m,1970}$ es el Producto Interno Bruto en 2002 del municipio m , $Q_{m,2000}$ es la producción petrolera en el año 2000 en el municipio m . y X'_m es el vector de controles geográficos y estructurales.

2. Intuición metodológica

¿Cuál es la intuición detrás de implementar Variables Instrumentales (VI) en vez de Mínimos Cuadrados Ordinarios (MCO)? En este escenario: Mencione al menos un sesgo del cual podríamos estar preocupados si interpretamos los efectos por MCO de manera causal. ¿Cuáles son los supuestos necesarios para que VI nos permita recuperar el efecto causal de interés? ¿Son creíbles los supuestos? Argumenten.

La razón principal para emplear el método de variable instrumental es que una estimación por Mínimos Cuadrados Ordinarios (MCO) podría arrojar resultados sesgados porque los ingresos municipales no son estrictamente exógenos. Municipios con poblaciones de mayores ingresos o con mayor capacidad institucional pueden mostrar al mismo tiempo mejores indicadores socioeconómicos, lo cual impediría identificar un efecto puramente causal. Existe también la posibilidad de causalidad inversa en la medida en que condiciones socioeconómicas más favorables permiten recaudar más ingresos fiscales. Además, no puede descartarse la influencia de factores no observados, como la calidad de las instituciones locales o la existencia de redes clientelares, que expliquen simultáneamente mayores ingresos y peores resultados en la provisión de bienes y servicios públicos.

Ante estos problemas los autores recurren a la producción de petróleo como instrumento de los ingresos municipales. La variación geográfica en la producción se interpreta como un shock exógeno que incrementa los recursos locales mediante regalías y permite recuperar el efecto causal de esos ingresos extraordinarios sobre el gasto público y los resultados sociales.

La validez de esta estrategia depende del cumplimiento de dos condiciones. La primera es la relevancia, que implica que la producción de petróleo Q_m esté correlacionada con los ingresos municipales R_m . Los autores muestran evidencia a favor de este punto al encontrar que un real adicional de producción petrolera se asocia con un incremento promedio de tres centavos en los ingresos municipales. La segunda es la exclusión, que exige que la producción de petróleo afecte los resultados socioeconómicos W_m únicamente a través de los ingresos municipales. Este supuesto se sostiene en el hecho de que los campos petroleros se ubican costa afuera, son operados por Petrobras y utilizan insumos y mano de obra altamente especializados que no provienen de los municipios beneficiarios, lo que reduce la posibilidad de un impacto directo sobre la provisión de servicios, las transferencias o los ingresos de los hogares.

La extracción de petróleo mar adentro limita de manera importante la existencia de derrames productivos locales y evita que los municipios reciban beneficios económicos distintos a las transferencias fiscales. Además, como las regalías se asignan según reglas legales vinculadas a la ubicación geográfica de los campos, el ingreso municipal depende de criterios externos y no de la capacidad de gestión local. Para reforzar este argumento, los autores presentan pruebas de falsificación que muestran que las condiciones socioeconómicas en 1970, antes de los descubrimientos offshore, no predicen la producción futura de petróleo. Esto sugiere que la asignación de los yacimientos no está asociada a características históricas de desarrollo municipal sino que se comporta de manera aleatoria una vez se controlan los factores geográficos.

En síntesis, los autores defienden de forma convincente que la localización costa afuera de los yacimientos impide la existencia de un canal productivo directo entre la actividad petrolera y el bienestar local, de modo que la única vía de transmisión entre la abundancia de petróleo y los resultados municipales es el aumento de regalías y transferencias fiscales.

3. Primera etapa

```
# Eliminar aleatoriamente el 10% de la base

# Número exacto de filas a eliminar (10% de la base)
eliminar <- round(0.1 * nrow(data))

# Escoger aleatoriamente los índices a eliminar
drop_index <- sample(1:nrow(data), size = eliminar)

# Crear la submuestra eliminando esos índices
data_final <- data[-drop_index, ]

# Solo municipios costeros que reciben petroleo
data_final <- subset(data_final, coastal == 1 & onshore == 0)

#Transformación de variables
data_final <- data_final %>%
  rename(
    revenue      = mun_budget_revenue2000_pred_c,
    oil_pc       = oilandgasvalue2000_cap,
    lat          = latitude1998,
    lon          = longitude1998,
    coast        = coastal,
    dist_fedcap  = dist_federal_capital1998,
    dist_statecap = dist_state_capital1998,
    state_cap    = state_capital,
    state        = sigla,
  )

# Cambiar a tipo factor
data_final <- data_final %>% mutate(state = as.factor(state))

# Primera etapa con FE de estado
prim_et <- lm(
  revenue ~ oil_pc + lat + lon + dist_fedcap + dist_statecap + state_cap + factor(state),
  data = data_final
)

#Resultados por cluster
vc <- vcovCL(prim_et, cluster = ~ state, type = "HC1")
ct <- coeftest(prim_et, vcov = vc)

# Solo coeficientes a reportar y sus SE
vars_keep <- c("oil_pc", "lat", "lon", "dist_fedcap", "dist_statecap", "state_cap")
idx <- match(vars_keep, rownames(ct))
betas1 <- ct[idx, "Estimate"]
ses1 <- ct[idx, "Std. Error"]

#Tabulación de los resultados
stargazer(
  prim_et,
```

```

type = "text",
coef = list(betas1),
se = ses1,
keep = paste0("^(", paste(vars_keep, collapse = "|"), ")$"),
dep.var.labels = "Ingresos municipales",
covariate.labels = c("Producción de petróleo",
                     "Latitud",
                     "Longitud",
                     "Dist. capital federal",
                     "Dist. capital estatal",
                     "Dummy capital estatal"),
omit = "factor\\(state_id\\)|\\(Intercept\\)",
star.cutoffs = c(.10, .05, .01),
digits = 3,
title = "Ingresos municipales sobre producción petrolera y controles"
)

```

En una tabla, presenten los resultados de estimar la primera etapa de la metodología de Variables Instrumentales. ¿Es el instrumento relevante y fuerte? Contesten esta pregunta usando los dos criterios recomendados en la clase magistral.

```

##
## Ingresos municipales sobre producción petrolera y controles
## =====
##                               Dependent variable:
##                               -----
##                               Ingresos municipales
## -----
## Producción de petróleo          0.035***
##
## Latitud                        55.510
##
## Longitud                       38.294
##
## Dist. capital federal          -0.282
##
## Dist. capital estatal          0.241
##
## Dummy capital estatal          342.259
##
## -----
## Observations                   141
## R2                           0.672
## Adjusted R2                   0.611
## Residual Std. Error           194.287 (df = 118)
## F Statistic                    10.983*** (df = 22; 118)
## =====
## Note:                         *p<0.1; **p<0.05; ***p<0.01

```

Los resultados de esta primera etapa muestran que la producción de petróleo es un instrumento relevante para explicar los ingresos municipales pues la relación entre estas variables es estadísticamente significativo incluso al

controlar por factores geográficos. Dado que se presenta un estadístico F de 10,9, el cual super el umbral de 10 recomendado, implica que el instrumento es lo suficientemente relevante. Este coeficiente estimado implica que por cada real per cápita adicional de producción petrolera, los municipios reciben en promedio 3,5 centavos de ingresos. Este resultado coincide con los hallazgos del artículo de investigación que interpretan la variación de la producción de petróleo offshore como un choque exógeno que incrementa los ingresos municipales a través de las regalías, reforzando la argumentación de su estrategia de identificación.

4. Segunda etapa

```
data_final <- data_final %>%
  rename(
    amc_code      = new_code_1970_1997,
    per_ho_elec   = prc_hhld_with_power_2000,
    salon_pob_2005 = clroomsMunicipal_pop2005,
    clinic_pob_2002 = estab_mun2002_without_pop,
    km_pob_2005    = km_paved_munic_c,
    ex_edu_cult    = pmun_exp_funct_educ_cult2000c,
    ex_salud_sani  = pmun_exp_funct_health_sanit2000c,
    ex_desarrollo_urb = pmun_exp_funct_hous_urban2000c,
    ex_transp      = pmun_exp_funct_transport2000c,
    ex_soc_transf  = pmun_exp_funct_welf2000c,
    perc_nofavelas = p_hhld_abovestandard_ppl_2000
  )
#Agrupación de variables de interés
outcomes <- c(
  "ex_edu_cult",
  "ex_salud_sani",
  "ex_desarrollo_urb",
  "ex_transp", "
  ex_soc_transf"
)

data_final$state <- as.factor(data_final$state)

#Estimación de la segunda etapa para cada una de las variables de resultado

m_edu_iv <- ivreg(ex_edu_cult ~ revenue + lat + lon + coast + dist_fedcap + dist_statecap + state_cap + fa
m_salud_iv <- ivreg(ex_salud_sani ~ revenue + lat + lon + coast + dist_fedcap + dist_statecap + state_cap + fa
m_urb_iv <- ivreg(ex_desarrollo_urb ~ revenue + lat + lon + coast + dist_fedcap + dist_statecap + state_c
m_trans_iv <- ivreg(ex_transp ~ revenue + lat + lon + coast + dist_fedcap + dist_statecap + state_cap + fa
m_soc_iv <- ivreg(ex_soc_transf ~ revenue + lat + lon + coast + dist_fedcap + dist_statecap + state_

#Errores agrupados por Estado
vc <- function(m) sandwich::vcovCL(m, cluster = ~ state, type = "HC1")
ses2 <- lapply(list(m_edu_iv,m_salud_iv,m_trans_iv,m_soc_iv), function(m) sqrt(diag(vc(m))))

#Presentación de resultados en tabla
```

```
stargazer(
  m_edu_iv, m_salud_iv, m_trans_iv, m_soc_iv,
  type = "text",
  se = ses2,
  dep.var.labels = c("Educación/Cultura",
                     "Salud/Saneamiento",
                     "Transporte",
                     "Asistencia Social"),
  covariate.labels = c("Ingresos (IV)",
                       "Latitud",
                       "Longitud",
                       "Costa",
                       "Dist. capital federal",
                       "Dist. capital estatal",
                       "Capital estatal"),
  omit = "factor\\(state\\)|\\(Intercept\\)",
  omit.labels = "Dummies de estado",
  star.cutoffs = c(.10, .05, .01),
  digits = 3,
  title = "Segunda etapa: ingresos instrumentados y gastos por función"
)
```

En una tabla, presenten e interpreten los resultados de estimar por VI el efecto de (mayores) ingresos públicos (producto del petróleo) sobre el gasto público reportado en el año 2000 en Educación y Cultura, Salud y Sanitización, Transporte y Transferencias Sociales.

```
##
## Segunda etapa: ingresos instrumentados y gastos por función
## =====
##                                     Dependent variable:
##                                     -----
##                                     Educación/Cultura Salud/Saneamiento Transporte Asistencia Social
##                                     (1) (2) (3) (4)
## -----
## Ingresos (IV)                0.139***      0.109***      0.123***      0.052***
##                               (0.006)      (0.005)      (0.005)      (0.004)
##
## Latitud                      4.708         -1.942         0.248         1.919
##                               (5.242)      (2.972)      (3.622)      (2.327)
##
## Longitud                     6.595         0.036         -6.448         0.623
##                               (5.730)      (3.036)      (4.159)      (1.572)
##
## Costa
##
## Dist. capital federal        -0.027         -0.090**      -0.049         0.012
##                               (0.048)      (0.042)      (0.045)      (0.026)
##
## Dist. capital estatal         0.033         0.032         0.002         0.022
##                               (0.055)      (0.028)      (0.039)      (0.022)
##
## Capital estatal             -6.907         17.773        -25.221*      53.113**
##                               (15.521)      (25.609)      (14.478)      (26.260)
##
## Constant                    -88.719         137.156       259.969*      -29.490
```

##	(186.855)	(108.802)	(157.680)	(50.762)
##				
## -----				
## Dummies de estado	Yes	Yes	Yes	Yes
## -----				
## Observations	141	141	141	141
## R2	0.750	0.715	0.521	0.604
## Adjusted R2	0.703	0.662	0.432	0.530
## Residual Std. Error (df = 118)	40.005	30.667	31.476	20.257
## =====				
## Note:			*p<0.1; **p<0.05; ***p<0.01	

Los resultados de la estimación mediante variables instrumentales confirman que un aumento en los ingresos municipales predichos por la producción de petróleo y gas se traduce en mayores niveles de gasto público en los distintos sectores analizados para el año 2000. Este hallazgo muestra una relación positiva y estadísticamente significativa entre los recursos fiscales extraordinarios y el gasto municipal, lo que respalda la hipótesis de que las regalías derivadas de la explotación petrolera tienen un impacto directo sobre la provisión de bienes y servicios públicos.

En el sector de Educación y Cultura, el coeficiente estimado es de 0.139, lo que implica que un incremento de una unidad en los ingresos municipales predichos se asocia con un aumento cercano a 0.14 unidades en el gasto destinado a este sector. Este resultado, significativo al 1%, evidencia que la inversión en capital humano constituye una prioridad en el uso de los recursos adicionales. En el ámbito de Salud y Saneamiento, el coeficiente de 0.109, también significativo, sugiere un patrón similar: los municipios con mayores ingresos derivados del petróleo tienden a asignar más recursos al fortalecimiento de la infraestructura y los servicios de salud, mejorando así la cobertura y la calidad en este campo.

El gasto en Transporte presenta un coeficiente de 0.123, significativo al 1%, lo que indica que los ingresos extraordinarios también se destinan con fuerza a la construcción y mantenimiento de infraestructura vial y de transporte. Por su parte, las Transferencias Sociales muestran un coeficiente positivo de 0.052, estadísticamente significativo aunque de menor magnitud en comparación con los otros sectores. Esto revela que, aunque los recursos petroleros se utilizan para apoyar programas de bienestar social, su peso relativo es menor frente a las áreas de educación, salud y transporte.

En conjunto, los resultados muestran que los choques positivos en ingresos municipales vinculados a la producción de petróleo offshore generan un aumento sustancial en el gasto público local, con mayor intensidad en los sectores de Educación, Salud y Transporte. Esta evidencia sugiere que los municipios costeros beneficiados canalizan los recursos principalmente hacia servicios públicos esenciales, mientras que las transferencias sociales reciben un refuerzo menor, lo que refleja una priorización clara en la asignación del gasto.

5. Forma reducida

```
# Personas que no viven en favelas en 2000
m_nofav <- lm(perc_nofavelas ~ oil_pc + lat + lon + coast + dist_fedcap + dist_statecap + state_cap + fact)

# Número de salones por millón de habitantes en el 2005.
m_classrooms <- lm(salon_pob_2005 ~ oil_pc + lat + lon + coast + dist_fedcap + dist_statecap + state_cap + fact)

# Errores estándar robustos agrupados por estado
se_nofav <- sqrt(diag(vcovCL(m_nofav, cluster = ~ state, type = "HC1")))
se_classrooms <- sqrt(diag(vcovCL(m_classrooms, cluster = ~ state, type = "HC1")))

#Tabulación forma reducida
stargazer(
  m_nofav, m_classrooms,
  type = "text",
```

```

se          = list(se_nofav, se_classrooms),
dep.var.labels = c("No-favelas", "Salones por millón de habitantes"),
covariate.labels = c("Producción de petróleo per cápita"),
keep        = "^oil_pc$",
omit        = "factor\\(state_id\\)|lat|lon|coast|dist_fedcap|dist_statecap|state_cap|Intercept",
omit.labels  = "Controles y efectos fijos",
star.cutoffs = c(.10,.05,.01),
digits      = 3,
title       = "Efecto del petróleo sobre indicadores de bienestar"
)

```

En una tabla, presenten e interpreten los resultados de estimar por VI el efecto de (mayores) ingresos públicos (producto del petróleo) sobre las variables efectivas de bienestar. En particular, sobre proporción de personas que no viven en favelas en 2000 y el número de salones por millón de habitantes en el 2005. ¿Qué indican sus resultados? ¿Mayores ingresos públicos resultan en niveles del bienestar medido a través de estas variables?. En texto describa los resultados para la primera variable .

```

##
## Efecto del petróleo sobre indicadores de bienestar
## =====
##                               Dependent variable:
##                               -----
##                               No-favelas Salones por millón de habitantes
##                               (1)          (2)
## -----
## Producción de petróleo per cápita -0.0005***          0.00000**
##                               (0.00004)          (0.00000)
## -----
## Controles y efectos fijos          Yes          Yes
## -----
## Observations          149          149
## R2          0.275          0.454
## Adjusted R2          0.148          0.358
## Residual Std. Error (df = 126)          5.015          0.002
## F Statistic (df = 22; 126)          2.169***          4.755***
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01

```

El coeficiente estimado de -0.0005 implica que, por cada aumento de una unidad en la producción de petróleo per cápita, la proporción de personas que no viven en favelas disminuye en 0.05 puntos porcentuales. Aunque el valor numérico del efecto puede parecer pequeño, el hecho de que sea altamente significativo estadísticamente al 1% indica que la relación es sistemática y no producto del azar.

Desde el punto de vista económico, este resultado es relevante porque muestra que los incrementos en la producción petrolera no se traducen en mejoras en las condiciones habitacionales, sino más bien en un ligero deterioro. Esto puede interpretarse como evidencia de que los ingresos del petróleo no se transforman de manera efectiva en políticas que reduzcan la informalidad en la vivienda. Por tanto, aun siendo cuantitativamente modesto, el hallazgo es importante porque cuestiona la eficacia con que los municipios canalizan los recursos petroleros hacia mejoras en bienestar habitacional al menos en esta variable de análisis.