

Tarea 2: Variables Instrumentales

Datos para la evaluación de impacto en políticas públicas Diplomado en Ciencia de Datos para Políticas Públicas

Pauta de corrección

Preguntas

A partir de la lectura del artículo, las clases del curso y los talleres, resuelva las siguientes preguntas.

1. ¿Por qué nos interesa estimar el impacto de los hijos en la participación laboral femenina? Reflexiones según el artículo y una perspectiva de política pública. **(4 puntos)**

No interesa saber el impacto causal que los hijos tienen sobre la participación laboral femenina. Esto nos ayuda a entender hasta que punto los hijos podrían explicar la baja participación de mujeres o el gender wage gap.

2. Señale cuál es el tratamiento, la variable resultado y los instrumentos de este estudio. **(4 puntos)**

En este caso tenemos dos tratamientos: 1) número de hijos que tiene la persona y 2) una variable binaria que identifica si la persona tiene más de dos hijos. A su vez, tenemos 5 variables de resultados, todos outcomes laborales: 1) binaria que identifica si la persona trabajó la semana anterior, 2) número de semanas trabajadas, 3) horas trabajadas a la semana, 4) ingresos relativos al trabajo y 5) logaritmo de los ingresos familiares. Finalmente, se cuenta con dos variables instrumentales: 1) si acaso los dos primeros hijos de una madre son del mismo sexo y 2) si acaso el segundo embarazo son gemelos.

3. Hay muchos estudios que muestran una correlación negativa entre fertilidad y oferta laboral para mujeres. ¿Por qué no le podemos atribuir causalidad a estos resultados? **(4 puntos)**

Hay bastante razones para creer que fertilidad y la participación laboral son decisiones conjuntas. En la práctica esto nos dice que es posible que las mujeres que no trabajan decidan tener hijos o que quienes tienen hijos decidan no trabajar. Va en ambas direcciones. También pueden explicar que las mujeres que tienen hijos son distintas a las que no los tienen, es posible que estas mujeres tengan, por ejemplo, un menor deseo de trabajar.

4. Los autores utilizan: 1) si acaso los dos primeros hijos de una madre son del mismo sexo y 2) si acaso el segundo embarazo son gemelos como instrumentos para el número de hijos. ¿Qué supuestos se deben cumplir para que estos instrumentos sean válidos? ¿Le parecen razonables estos supuestos? **(4 puntos)**

Se deben cumplir los 5 supuestos de variables instrumentales:

1. SUTVA: que el que una mujer tenga un hijo no afecte el outcome de otra mujer (basicamente que no hayan externalidades)
2. Independencia: las mujeres que tienen hijos de igual o distinto sexo son comparables y las que tienen gemelos son comparables a las que no los tienen.
3. Exclusión: tener dos hijos del mismo sexo o gemelos solo afecta mis resultados via afectar mi probabilidad de tener un tercer hijo
4. First Stage: Ambos instrumentos afectan la probabilidad de tener un tercer hijo
5. Monotonicidad: Ambos instrumentos sólo aumentan mi probabilidad de tener un tercer hijo.

En términos generales, los supuestos son razonables. Una potencial amenaza al supuesto de independencia en el caso del primer instrumento es que en países donde hay aborto selectivo, el sexo del segundo hijo podría no ser realmente aleatorio. Otra potencial amenaza es que el instrumento de gemelos podría violar el supuesto

de exclusión si el efecto de tener gemelos va más allá del efecto en la fertilidad. Es posible que la carga que supone criar gemelos obligue a una madre a abandonar el mercado laboral, pero que la misma madre logre seguir trabajando si tiene dos hijos distanciados en el tiempo. En el sentido contrario, una madre que tiene mellizos podría tener menos interrupciones en su trayectoria laboral (menos embarazos y partos) que una madre que tiene dos hijos distanciados en el tiempo, lo que podría traducirse en una mayor adhesión al mercado laboral.

5. ¿Cuál es el objetivo de la Tabla 4? ¿Qué concluyen los autores a partir de estos resultados? **(3 puntos)**

La tabla 4 corresponde a un balance entre características observadas. En otras palabras, estima las diferencias de medias en una serie de variables de interés entre mujeres que tienen hijos del mismo sexo versus las que no tienen hijos del mismo sexo. Esto se aplica en la columna 1 y 2 para los resultados de la encuesta en 1980 y 1990. La tercera columna estima el balance para las mujeres que tienen gemelos versus las que no tienen gemelos según los datos de la encuesta de 1980. En general podemos notar algunas diferencias menores entre los grupos, sobretodo para el grupo de personas que tienen gemelos versus el resto (puede ser más preciso en esta lectura).

6. Replique e interprete los resultados de la Tabla 5, Columna 1, filas 1 a 2 usando una regresión lineal. **(10 puntos)**

```
# Vamos a cargar las librerías de trabajo
library(rio)
library(dplyr)
library(kableExtra)
library(texreg)
library(estimatr)
library(AER)

# Importamos la BBDD
# data <- haven::read_dta("ang_ev_1980.dta")
data_iv <- import("ang_ev_1980.dta")

## Estimemos el modelo solicitado
m1 <- lm(morekids ~ samesex, data=data_iv)
m2 <- lm(kidcount ~ samesex, data=data_iv)

texreg(l=list(m1, m2), include.ci = FALSE, digits = 4,
       caption="First Stage",
       custom.header = list("Tratamiento \\vspace{0.2 cm}" = 1:2),
       custom.model.names = (c("Más de dos hijos", "Número de hijos")),
       custom.coef.names=c("Intercepto", "Instrumento"),
       float.pos="h", single.row = FALSE, fsingle.row = TRUE,
       caption.above = TRUE, stars = c(0.01, 0.05, 0.1),
       custom.note = ("\\parbox{.6\\linewidth}{\\vspace{4pt}%stars.
                       Error estándar entre paréntesis.}"))
```

Table 1: First Stage

	Tratamiento	
	Más de dos hijos	Número de hijos
Intercepto	0.3751*** (0.0011)	2.5211*** (0.0018)
Instrumento	0.0595*** (0.0015)	0.0759*** (0.0026)
R ²	0.0037	0.0022
Adj. R ²	0.0037	0.0022
Num. obs.	402014	402014

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Error estándar entre paréntesis.

7. ¿Qué porcentaje de mujeres con hijos de distinto sexo tienen un 3er hijo? ¿Qué porcentaje de mujeres con hijos de un mismo sexo tienen un tercer hijo? ¿Cual es la tasa de compliers? **(3 puntos)**

Un 37,5% de mujeres con hijos de distinto sexo tienen un tercer hijo. Un (37,5+6) 43,5% de mujeres con hijos del mismo sexo tienen un tercer hijo. La tasa de compliance es 6%.

8. Replique los resultados de la Tabla 5, Columna 2, filas 3 a 4 usando 2SLS **(6 puntos)**

```
# Tabla 5 columna 2 filas 3 y 4
m23 <- ivreg(mom_worked ~ morekids | samesex, data = data_iv)
m24 <- ivreg(mom_weeks_worked ~ morekids | samesex, data = data_iv)
texreg(l=list(m23,m24), include.ci = FALSE, digits = 4,
       float.pos="h", single.row = FALSE, fsingle.row = TRUE,
       caption.above = TRUE, stars = c(0.01, 0.05, 0.1))
```

Table 2: Statistical models

	Model 1	Model 2
(Intercept)	0.6176*** (0.0106)	23.3100*** (0.4758)
morekids	-0.1286*** (0.0261)	-6.1372*** (1.1713)
R ²	0.0142	0.0173
Adj. R ²	0.0142	0.0173
Num. obs.	402014	402014

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

9. Interprete estos resultados. ¿Es este el ATE de tener más de 2 hijos? **(3 puntos)**

No, este es el LATE, es el efecto de tener un tercer hijo en participación laboral en mujeres que decidieron tener un tercer hijo porque sus dos primeros hijos eran del mismo sexo, pero que de haber tenido dos hijos del mismo sexo no habrían tenido un tercero.

10. Replique e interprete los resultados de la Tabla 5, Columna 7, filas 1 a 4 usando una regression lineal. ¿Qué porcentaje de mujeres cuyo segundo embarazo son gemelos tienen un tercer hijo? ¿Cual es la tasa de compliers? **(6 puntos)**

```
m71 <- lm(morekids ~ twins_2, data=data_iv)
m72 <- lm(kidcount ~ twins_2, data=data_iv)
m73 <- lm(mom_worked ~ twins_2, data=data_iv)
```

```
m74 <- lm(mom_weeks_worked ~ twins_2, data=data_iv)

texreg(l=list(m71, m72, m73, m74), include.ci = FALSE, digits = 4,
       float.pos="h", single.row = FALSE, fsingle.row = TRUE,
       caption.above = TRUE, stars = c(0.01, 0.05, 0.1))
```

Table 3: Statistical models

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.4000*** (0.0008)	2.5526*** (0.0013)	0.5658*** (0.0008)	20.8408*** (0.0353)
twins_2	0.6000*** (0.0084)	0.8062*** (0.0140)	-0.0460*** (0.0085)	-2.0028*** (0.3823)
R ²	0.0126	0.0082	0.0001	0.0001
Adj. R ²	0.0126	0.0082	0.0001	0.0001
Num. obs.	402014	402014	402014	402014

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Un 100% de las mujeres cuyo segundo embarazo son gemelos tienen un tercer hijo. La tasa de complaine es 60%.

11. Replique e interprete los resultados de la Tabla 5, Columna 8, filas 3 a 4 usando 2SLS (10 puntos)

```
m83 <- ivreg(mom_worked ~ morekids | twins_2, data = data_iv)
m84 <- ivreg(mom_weeks_worked ~ morekids | twins_2, data = data_iv)

texreg(l=list(m83, m84), include.ci = FALSE, digits = 4,
       float.pos="h", single.row = FALSE, fsingle.row = TRUE,
       caption.above = TRUE, stars = c(0.01, 0.05, 0.1))
```

Table 4: Statistical models

	Model 1	Model 2
(Intercept)	0.5965*** (0.0058)	22.1760*** (0.2587)
morekids	-0.0767*** (0.0141)	-3.3380*** (0.6327)
R ²	0.0124	0.0140
Adj. R ²	0.0124	0.0140
Num. obs.	402014	402014

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

12. ¿Porqué podrían los efectos estimados utilizando `samesex` versus `twins_2` como instrumento ser diferentes? (3 puntos)

Podrían ser diferentes porque en un case estamos estimando el efecto de etenr un tercer hijo en mujeres que tuvieron un tercer hijo porque sus dos primeros hijo fueron del mismo sexo. En el segundo caso es el efecto de etner un tercer hijo en mujeres que tuvieron un tercer hijo por tener gemelos. Las poblaciones son distintas y el efecto tratamiento podría ser distinto en estos dos grupos.