

**Taller 1**  
**Econometría Avanzada, 2022-2**  
**Profesor: Manuel Fernández**

**Carlos A. Ayala – 201911488**

*Todos los elementos utilizados para la solución de este taller (bases de datos, dofile, outputs y demás), pueden ser consultados en el siguiente enlace:*  
[https://github.com/Mateocontrerass/Taller\\_2\\_Avanzada](https://github.com/Mateocontrerass/Taller_2_Avanzada)

**1. Primer ejercicio:**

El gobierno ha realizado un seguimiento a la condición de desnutrición de los niños de colegios durante los últimos años. Durante este tiempo, ha observado que la desnutrición en los niños presenta un problema de gran relevancia que es necesario atacar. Por ende, el gobierno ha lanzado un programa de nutrición enfocado a dar clase a los padres de los estudiantes en cuestiones referentes a una buena alimentación. Así las cosas, se usa la base ENSIN, para la cual se hacen 5 divisiones en subgrupos con diferentes tratamientos.

**1. Subgrupo 1:** En este grupo los estudiantes fueron asignados de forma aleatoria a las capacitaciones nutricionales. Sin embargo, se hizo la asignación en dos etapas, las cuales usaron aleatorizaron en diferentes niveles:

- Etapa 1: Nivel colegio. Se capacitaron a todos los estudiantes de colegios aleatoriamente seleccionados.
- Etapa 2: Nivel estudiante. Se realizó una asignación aleatoria estratificada por colegio, es decir se seleccionaron aleatoriamente a colegios y posteriormente se realizó una asignación aleatoria de los ´ estudiantes dentro de los colegios seleccionados.

**2. Subgrupo 2:** A colegios elegidos de forma aleatoria se les otorgo una capacitación sobre bullying, mientras que a los otros no se les asigno ningún tipo de capacitación.

**3. Subgrupo 3:** En este grupo todos los colegios recibirán la capacitación, pero con distintos grados de información. Por un lado, de forma aleatoria, a algunos de ellos se les informa desde el primer momento de la capacitación que serán evaluados con un test final. Por otro lado, aunque los otros colegios también reciben la capacitación, en ellos no se realiza ningún tipo de test de conocimientos y los padres son informados de ello.

**4. Subgrupo 4:** De forma aleatoria se seleccionan algunos colegios, los cuales ninguno recibirá la capacitación. Sin embargo, a algunos de ellos se les informa de su estatus de tratamiento, es decir, se les informa a los padres que otros colegios recibieron la capacitación nutricional, pero ellos no. A los otros colegios de este grupo no se les comenta acerca de su estatus como grupo de control, de manera que no saben que no se encuentran recibiendo la capacitación ni de la existencia del programa.

**5. Subgrupo 5:** De forma aleatoria se seleccionan colegios donde se les comentaran mensajes distintos a los padres acerca de la importancia de la nutrición: a algunos padres se les dará la información A: “Aunque varios estudios muestran que la alimentación es clave para la formación del niño, otros estudios muestran que otros factores como el número de profesores y la calidad del colegio son aún más relevantes.” En los demás colegios, los padres recibieran la información B: “Una deficiencia alimenticia a la edad preescolar, reduce en forma permanente la destreza de los niños.”

Tabla 1: Información de los grupos de aleatorización

Muestra	Submuestra	% de la muestra	Grupo de efecto	Etapas/nivel de aleatorización	Submuestra de etapas	% de la submuestra de etapas
50639 [1649]	23039 [542]	45.49 % [32.82 %]	Subgrupo 1	Etapa 1 Etapa 2	15388 [402] 7651 [140]	66.79 % [74.16 %] 33.20 % [25.83 %]
	6900 [277]	13.62 % [16.79 %]				
	6900 [277]	13.62 % [16.79 %]	Subgrupo 2	Nivel colegio		
	6900 [277]	13.62 % [16.79 %]	Subgrupo 3	Nivel colegio		
	6900 [277]	13.62 % [16.79 %]	Subgrupo 4	Nivel colegio		
	6900 [276]	13.62 % [16.73 %]	Subgrupo 5	Nivel colegio		

■ Nota: Entre corchetes se encuentra la información de nivel colegio y sin corchetes representa la información a nivel estudiante. Esto es clave para entender e interpretar los diferentes niveles de aleatorización.

**a. Responda las siguientes preguntas teniendo en cuenta el contexto y los datos disponibles.**

- i. Proponga uno o varios modelos de regresión que recupere el efecto de recibir el tratamiento sobre el tallaje. Describa las variables, los parámetros, el o los subgrupos a evaluar, y describa los supuestos para que el estimador sea consistente, y si los supuestos de identificación son plausibles teóricamente.**

Antes de proponer un modelo es necesario establecer el subgrupo sobre el cual se va a realizar el estudio. Se proponen 4 subgrupos con diferentes especificaciones de tratamiento, dentro de los cuales es el subgrupo 1 el que presenta el tratamiento adecuado dadas las

características del programa implementado por el gobierno. Ahora, para que el análisis estimado de la política capture un efecto causal es necesario que se cumplan 2 supuestos principales. El primero de ellos es que la aplicación del tratamiento sea independiente de los resultados potenciales de los grupos de tratamiento y control. Esto asegura que no existen diferencias previas entre los grupos de control y tratamiento que expliquen los resultados obtenidos. Llamado de otra forma, se requiere que los grupos estén balanceados en sus características. En segundo lugar, se requiere que no exista contaminación de las observaciones por los resultados potenciales de otras observaciones. Este supuesto, SUTVA, es muy importante puesto que denota que existe independencia de los resultados de una observación  $i$  de los resultados de una observación  $j$ . Dada la estructura del estudio, se requiere que las unidades del grupo de control no se vean afectadas por el tratamiento.

Bajo los supuestos anteriores, el modelo que capturaría el efecto causal de la política sobre el tallaje de los niños es el siguiente:

$$Tallaje_i = \beta_0 + \beta_1 D_i + \beta_2 Educación_i + \beta_3 Ingreso_i + \beta_4 Ocupado_i + \beta_5 Raza_i + \beta_6 Sexo_i + \beta_7 Personas_i + e_i$$

- Tallaje: Medidas de tallaje de los niños. Variable a explicar.
- D: Variable dummy que toma valor de 1 si el estudiante fue tratado o no. Bajo los supuestos planteados, el coeficiente asociado captura el efecto causal de la política sobre el tallaje de los estudiantes.
- Educación jefe de hogar: Número de años de educación formal del jefe de hogar. El coeficiente asociado captura el efecto de un año adicional de educación del jefe de hogar sobre el tallaje de los estudiantes.
- Ingreso jefe de hogar: Ingreso mensual en miles de pesos del jefe de hogar. El coeficiente asociado captura el efecto de mil pesos adicionales en ingreso mensual del jefe de hogar sobre el tallaje de los estudiantes.
- Ocupado: Variable dummy que toma valor de 1 si el jefe de hogar está ocupado. El coeficiente asociado captura el efecto de ser ocupado, respecto a no estarlo, sobre el tallaje de los estudiantes.
- Raza: Variable dummy que toma valor de 1 si pertenece a raza afro o indígena. El coeficiente asociado captura el efecto de ser afro o indígena, respecto a no serlo, sobre el tallaje de los estudiantes.
- Sexo: Variable dummy que toma el valor de 1 si es hombre. El coeficiente asociado captura el efecto de ser hombre, respecto a ser mujer, sobre el tallaje de los estudiantes.
- Personas: Número de personas que viven en el hogar. El coeficiente asociado captura el efecto de una persona adicional que viva en el hogar sobre el tallaje de los estudiantes.

Finalmente, es necesario mencionar que los supuestos necesarios para que el modelo planteado capture efectos causales no se cumplen en este caso. La aleatorización para la asignación del tratamiento se realizó en 2 etapas, donde la primera etapa aleatoriza entre colegios y todos los estudiantes de los colegios seleccionados reciben el tratamiento. Sin embargo, la segunda etapa aleatoriza adicionalmente entre estudiantes, haciendo que solo un subgrupo de estudiantes de los colegios seleccionados reciba el tratamiento. De esta forma, el supuesto SUTVA no puede mantenerse sobre toda la muestra y el coeficiente estimado no captura el efecto causal de la política.

- ii. **Suponga que los supuestos se cumplen. Estime el modelo, presente los resultados en una tabla, e interprete los resultados y su significancia estadística.**

VARIABLES	(1) Talla
Grupo_1	0.255*** (0.0189)
Educacion_jefe	0.0201*** (0.00222)
Ingreso_jefe	0.000592*** (0.000101)
Ocupado_jefe	-0.0538*** (0.0204)
Raza_afro_indig	-0.424*** (0.0164)
Sexo	-0.00635 (0.0164)
Personas_hogar	-0.0488*** (0.00425)
Constant	-0.364*** (0.0360)
Observations	23,039
R-squared	0.049

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Si se pudiera mantener los supuestos planteados, el efecto estimado indica que haber sido tratado incrementa la talla del estudiante en 25,5%. Bajo los supuestos, esta interpretación correspondería al efecto causal promedio del tratamiento, y se concluiría que la política tuvo efectos positivos sobre la desnutrición de los niños.

**b. Realice una tabla**

VARIABLES	(1) Grupo_1	(2) Grupo_1_E1	(3) Grupo_1_E2
Educacion_jefe	0.00134* (0.000778)	0.00121 (0.000955)	0.00154 (0.00134)
Ingreso_jefe	-5.63e-05 (3.51e-05)	-9.79e-05** (4.20e-05)	2.77e-05 (6.30e-05)
Ocupado_jefe	0.00372 (0.00709)	0.00298 (0.00871)	0.00507 (0.0122)
Personas_hogar	0.000646 (0.00149)	0.000780 (0.00184)	0.000323 (0.00255)
Raza_afro_indig	0.00252 (0.00572)	0.00481 (0.00703)	-0.00247 (0.00984)
Sexo	0.00482 (0.00572)	0.00734 (0.00703)	-7.35e-05 (0.00983)
Constant	0.237*** (0.0123)	0.242*** (0.0152)	0.228*** (0.0210)
Observations	23,039	15,388	7,651
R-squared	0.000	0.001	0.000

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Como se puede observar, los 3 subgrupos presentan signos de una buena aleatorización. Se hace uso del concepto partialling out para establecer si existe una correcta aleatorización. Se observa que para la asignación en el subgrupo 1 la educación del jefe de hogar es significativa. Así mismo, se observa que el ingreso del jefe de hogar es significativo. Esto puede ser señal de una posible diferencia previa al tratamiento; no obstante, la significancia no es lo suficientemente fuerte, así como el valor del coeficiente es muy pequeño, lo que seguramente no representa una diferencia significativa en los grupos de tratamiento y control y se concluye que existe una relativamente buena aleatorización.

**c. Identifique si se cumplen los supuestos SUTVA**

**i. Establezca una relación matemática del cumplimiento de SUTVA**

Para que se cumpla el supuesto de SUTVA, debe cumplirse que los resultados potenciales de una unidad no estén correlacionados con el estado de tratamiento de otras unidades. Matemáticamente:

$$E[y_i|D_i, D_{-i}] = E[y_i|D_i] + E[y_i|D_{-i}] = E[y_i|D_i]$$

En este caso, se sospecha la existencia de contaminación en el grupo de control al interior del subgrupo 1 en la 2 etapa de aleatorización. Esto debido a que este puede verse afectado por el tratamiento como resultado de la convivencia de unidades tratadas y no tratadas en un mismo ambiente (comparten colegio, y posiblemente sean estudiantes de un mismo curso).

**ii. Describa un proceso para corroborar la existencia de SUTVA**

Para corroborar que se cumple el supuesto, se propone comparar los resultados potenciales de haber sido asignado al tratamiento en la etapa 1 respecto a los resultados potenciales de haber sido asignado al tratamiento en la etapa 2. La lógica está en que, si se cumple SUTVA en las 2 unidades, los resultados potenciales de no recibir el tratamiento deberían ser los mismos independientemente de la etapa de asignación. Para esto, es necesario realizar una tabla de balanceo muestral con el objetivo de corroborar que no existen diferencias previas que puedan afectar el resultado potencial, y que este solo se afecta por la etapa en la que fue asignado el tratamiento. Para ello, se crea una variable dummy (E1) que es igual a 1 si la observación pertenece al grupo de control de la etapa 1, y 0 si pertenece al grupo de control de la etapa 2.

**iii. Realice el proceso descrito**

A continuación, se muestra la tabla de balanceo muestral.

Tabla de Balanceo Muestral E1				
	Y1_mean	Y0_mean	Y1-Y0	p-value
Educacion	6.3810	6.3806	0.0004	0.9955
Ingreso	74.2711	73.6929	0.5782	0.6680
Ocupado	0.7962	0.7968	-0.0006	0.9217
Personas	5.1338	5.1532	-0.0194	0.5565
Raza	0.5051	0.4923	0.0128	0.1113
Sexo	0.5143	0.5158	-0.0015	0.8505

Tabla 1.4

La tabla de balanceo muestral evidencia que no existen diferencias previas en los controles entre el grupo de asignación de la etapa 1 y la etapa 2, y que por consiguiente el efecto promedio del tratamiento debería ser el mismo en ambos grupos bajo el supuesto de SUTVA puesto que en ambos grupos se hizo una buena aleatorización. Ahora, basta con mostrar una diferencia de medias para poder corroborar si se cumple el supuesto o no. A continuación, se muestran los resultados.

Diferencia de medias para E1				
	Y1_mean	Y0_mean	Y1-Y0	p-value
Talla	-0.7229	-0.6605	-0.0625	0.0023

Tabla 1.5

La tabla muestra que existe efectivamente diferencia en el resultado promedio de los grupos de control, en donde el grupo de control de la etapa 1 presenta una media menor que el grupo de control de la etapa 2. Esto corrobora la sospecha de contaminación para el grupo de control de la etapa 2, el cual resulta contaminado al convivir en un mismo ambiente con unidades tratadas.

#### d. Pruebe la existencia de sesgos comportamentales.

Para probar los diferentes sesgos se hace uso de los 4 subgrupos adicionales que ofrece la base de datos. Estos subgrupos contienen tratamientos que funcionan acorde a lo esperado para probar los diferentes sesgos especificados.

Así, el subgrupo 2 sirve para probar el sesgo por efecto placebo, el cual gracias a la existencia de un tratamiento que en principio no debería afectar el tallaje de los niños permite estudiar la existencia de este efecto comportamental. El subgrupo 3 sirve para probar la existencia de sesgo por efecto Hawthorne gracias al componente de evaluación presente en el grupo de tratamiento. Esta situación se compara con observar a las unidades tratadas, y plantea una buena situación para estudiar dicho efecto comportamental. El subgrupo 4, donde ninguno de los grupos es tratado, permite estudiar el efecto John Henry puesto que a uno de los dos grupos se le da información de su estado de tratamiento y al otro no. Finalmente, el subgrupo 5 permite estudiar el efecto de demanda experimental pues, si bien no se otorga el mismo tipo de tratamiento que al subgrupo 1, se brinda información sugestiva respecto al experimento, lo que permite estudiar si los individuos pueden intentar predecir el objetivo del experimento y cambiar su comportamiento al respecto. Estudiar la existencia de los efectos mencionados se reduce a realizar una regresión de la variable de interés respecto a ser asignado al tratamiento de los diferentes subgrupos, y estudiar el signo del coeficiente. A continuación, se presenta los resultados de la regresión.

VARIABLES	(Ef. Placebo) Talla	(E. Hawthorne) Talla	(E. John Henry) Talla	(E. Demanda Experimental) Talla
-----------	------------------------	-------------------------	--------------------------	------------------------------------

Grupo_2	-0.0445 (0.0310)			
Grupo_3		0.0259 (0.0311)		
Grupo_4			0.0609* (0.0314)	
Grupo_5				-0.0643** (0.0311)
Educacion_jefe	0.0184*** (0.00405)	0.0158*** (0.00409)	0.0157*** (0.00409)	0.0157*** (0.00409)
Ingreso_jefe	0.000640*** (0.000190)	0.000449** (0.000189)	0.000444** (0.000189)	0.000440** (0.000189)
Ocupado_jefe	-0.0532 (0.0372)	-0.0162 (0.0379)	-0.0142 (0.0379)	-0.0148 (0.0379)
Raza_afro_indig	-0.475*** (0.0299)	-0.395*** (0.0301)	-0.398*** (0.0301)	-0.396*** (0.0301)
Sexo	-0.0173 (0.0299)	0.00815 (0.0301)	0.00383 (0.0301)	0.00710 (0.0301)
Personas_hogar	-0.0542*** (0.00774)	-0.0415*** (0.00769)	-0.0415*** (0.00769)	-0.0412*** (0.00769)
Constant	-0.177*** (0.0684)	-0.445*** (0.0664)	-0.458*** (0.0684)	-0.407*** (0.0679)
Observations	6,900	6,900	6,900	6,900
R-squared	0.050	0.034	0.035	0.035

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

- i. **Efecto Placebo:** Matemáticamente, se puede expresar así. Sea  $\tau = E[Y|D = 1] - E[Y|D = 0]$ , si  $D$  es un placebo,  $E[Y|D = 1] - E[Y|D = 0] = 0$ . Así, si  $D$  es un placebo, no debería existir diferencia en los resultados potenciales del ser tratado respecto a no serlo. En este caso, se observa que en el subgrupo 2, el grupo que recibió el tratamiento disminuyó su tallaje en 4,5% respecto al grupo de control. Esto significa que recibir charlas sobre el bullying no tiene efectos positivos sobre el tallaje de los estudiantes, y por consiguiente se corrobora la no existencia de efectos placebo.
- ii. **Efecto Hawthorne:** Matemáticamente, se puede expresar así. Sea  $\tau = E[Y|D = 1] - E[Y|D = 0]$ , donde  $D$  representa ser observado dentro del tratamiento,  $E[Y|D = 1] - E[Y|D = 0] > 0$ . Así, si se observan a las unidades tratadas, estas definitivamente deberían presentar un mayor resultado potenciales que aquellas unidades no observadas. En este caso, se observa que al establecer una evaluación al final de la capacitación el tallaje incrementa en 2,6%. Esto refleja que los individuos mejoran su desempeño bajo situaciones



donde se sienten observados, y por lo tanto existe evidencia del efecto Hawthorne.

- iii. **Efecto John Henry:** Matemáticamente, se puede expresar así. Sea  $\tau = E[Y|D = 1] - E[Y|D = 0]$ , donde D representa tener conocimiento del estado como unidad de control,  $E[Y|D = 1] - E[Y|D = 0] > 0$ . Así, las unidades que saben que son grupo de control deberían presentar un mayor resultado potenciales que aquellas unidades que no lo saben. En este caso, En este caso se observa que al mencionarle al grupo de control su estado de tratamiento, su tallaje incrementa en 6% respecto al grupo donde no se menciona su estado. Así, se evidencia la existencia de un cambio en el comportamiento para compensar su estado de no tratado. Por lo tanto, se corrobora la existencia del efecto John Henry.
- iv. **Efecto Demanda Experimental:** Matemáticamente, se puede expresar así. Sea  $\tau = E[Y|D = 1] - E[Y|D = 0]$ , donde D representa tener información respecto al objetivo del tratamiento,  $E[Y|D = 1] - E[Y|D = 0] > 0$ . Así, las unidades que tienen información respecto al objetivo del estudio deberían presentar un mayor resultado potencial que aquellas que no. En este caso, el grupo que se le da información respecto a la importancia de la buena alimentación en la destreza de los niños, pero no recibe la cátedra que se otorga al subgrupo 1, disminuyó el tallaje de los niños en 6,5%. Esto muestra que los individuos no cambian su comportamiento al intentar predecir el objetivo del tratamiento, o por lo menos que el grupo no logra predecir el objetivo del estudio. Por tal motivo, no puede corroborarse la existencia del efecto de Demanda Experimental.

**e. Determine, con base a los incisos b, c y d, si la estimación del inciso a recupera un efecto causal. Si no es afirmativo, ¿Cómo modificaría el experimento?**

En los puntos anteriores se probó que en el subgrupo 1 existió una correcta aleatorización. De igual forma, se probó que el supuesto de SUTVA solo se cumplía en la etapa 1 y no en la etapa 2, debido a coexistencia de unidades de tratamiento y de control. Finalmente se probó la posible existencia de efecto Hawthorne y efecto John Henry en el experimento. En esta última parte, se debe hacer distinción en la existencia de los efectos mencionados. En la etapa 1, debido a la forma en cómo se construyó el experimento, no es posible la existencia del efecto Hawthorne debido a que no se evalúa el desempeño de los estudiantes al terminar las capacitaciones. La misma situación se plantea en la etapa 2, la cual tampoco presenta una evaluación final que suponga la existencia del efecto Hawthorne. No obstante, el panorama cambia cuando se analiza la existencia del efecto John Henry. En la etapa 1 es posible suponer que no hay forma que exista el efecto por construcción del experimento, todas las unidades al interior de un colegio tratado son tratadas. Al contrario, la coexistencia de unidades tratadas y no tratadas en la etapa 2 supone que las unidades de control pueden descubrir su estado de tratamiento al cruzar información con las unidades

tratadas. Así, las unidades de control pueden cambiar su comportamiento, y efectivamente se observa que así sucede. Se corrobora la existencia de este efecto por medio de la diferencia de medias efectuada en el punto c, donde la media del tallaje en el grupo de control de la etapa 1 es menor a la media del grupo de control de la etapa 2.

Por lo anterior, se establece que no se cumple SUTVA al interior de la etapa 2, así como la existencia del efecto John Henry. Por tal motivo, el modelo planteado en el inciso a no recupera el efecto causal del programa al incluir la etapa 2 dentro de las unidades de observación. El experimento se puede modificar de dos formas. La primera, sería restringir la muestra únicamente a la etapa 1 y efectuar allí el modelo planteado en el inciso 1. Bajo esas condiciones sería posible capturar un efecto causal del tipo LATE, el cual mediría el efecto causal local al interior de la etapa 1. La segunda modificación sería establecer una única forma de aleatorización al interior del subgrupo 1, la etapa 1, la cual permita que al interior de los colegios solo existan observaciones con un mismo estado de tratamiento. Bajo estas condiciones, sería posible implementar el modelo planteado en el inciso a, y el resultado capturaría el efecto causal del tipo ATE. Debido a que las 2 modificaciones son similares, pero la modificación 2 permite encontrar el ATE, la modificación 2 sería elegida.

## 2. Segundo ejercicio:

Card (2009) estudia la relación que existe entre la migración y la desigualdad salarial. Para ello, usa los flujos de migración a Estados Unidos desde diferentes países del mundo. A relación Card explora en su artículo es la siguiente.

$$w_{lj} = \beta_0 + \tau \ln(r_{lj}) + \beta_2 X_l + e_{lj}$$

**a. Explique qué tipo de relación captura  $\tau$  entre  $w_{lj}$  y  $r_{lj}$ . ¿cree que el estimador de MCO es consistente?**

El parámetro  $\tau$  captura el efecto de la cantidad de horas trabajadas de los extranjeros del país  $j$  respecto a los nativos sobre la brecha salarial de los extranjeros del país  $j$  respecto a los nativos. En este caso, el efecto que captura  $\tau$  no es consistente puesto que  $r_{lj}$ , si bien hace distinción entre el tipo de educación de los migrantes, no se garantiza que no existan condiciones adicionales que hayan influenciado en la cantidad de horas trabajadas. Así las cosas, existen condiciones adicionales como la disposición a trabajar, o prejuicios por parte de los empleadores que fuercen el tiempo de trabajo de los inmigrantes a incrementar o disminuir, y repercutir en el salario. Por lo tanto, el estimador no será consistente si no se toma en consideración este efecto.

Preocupado por problemas de endogeneidad, Card propone un instrumento tipo Bartik. El instrumento es de la forma

$$B_{lj} = \sum_c z_{lc,1980} g_{cj}$$

Donde

$$z_{lc,1980} = \frac{N_{lc,1980}}{N_{c,1980}} * \frac{1}{P_{l,2000}}$$

**b. Con base en la estrategia de Card**

- i. **Explique por qué el instrumento tipo Barik puede predecir cambios en los ratios de horas trabajadas entre migrantes y nativos.**

Tal como está planteado, el instrumento tipo Bartik planteado puede predecir los cambios en los ratios de horas trabajadas entre inmigrantes y nativos debido a que establece un marco de referencia del ancho de mercado que acapara cada población migrante. Dicho de otra forma, instrumenta el tamaño del mercado propio a cada población migrante, lo cual explica la cantidad de horas trabajadas debido a la cantidad de migrantes de un país  $c$  existentes en la ciudad  $l$ . Entre mayor sea la cantidad de migrantes del país  $c$  en la ciudad  $l$ , mayor será la cantidad de horas trabajadas por esa población específica en la ciudad  $l$ .

- ii. **Plantee la ecuación estructural, la primera etapa y la forma reducida.**

**Ecuación Estructural:**  $w_{lj} = \beta_0 + \tau \ln(r_{lj}) + \beta X + e_{lj}$ , donde  $X$  es el vector de características de la ciudad (controles).

**Primera Etapa:**  $\ln(r_{lj}) = \alpha_0 + \alpha_1 B_{lj} + \alpha X + \eta_{lj}$

**Forma reducida:**  $w_{lj} = \beta_0 + \tau(\alpha_0 + \alpha_1 B_{lj} + \alpha X + \eta_{lj}) + \beta X + e_{lj}$

$$w_{lj} = \beta_0 + \tau\alpha_0 + \tau\alpha_1 B_{lj} + \tau\alpha X + \beta X + \tau\eta_{lj} + e_{lj}$$

$$w_{lj} = \gamma_0 + \phi B_{lj} + \gamma X + \mu_{lj}$$

Donde  $\gamma_0 = \beta_0 + \tau\alpha_0$ ,  $\phi = \tau\alpha_1$ ,  $\gamma = \tau\alpha + \beta$  y  $\mu_{lj} = \tau\eta_{lj} + e_{lj}$

El supuesto de relevancia establece que, condicional en  $X_l$ , debe existir al menos un  $k$  para el cual  $z_{kl,1980}$  tenga poder predictivo sobre  $\ln(r_{lj})$  y que esta influencia sea tal que no se cancele al agregar los choques. Por su parte, exogeneidad estricta es la restricción de exclusión tradicional, pero esta recae sobre los “shifts”. Esencialmente, lo que esta condición plantea es que las poblaciones de inmigrantes en 1980 provenientes de un país  $c$ ,

condicional en  $X_l$ , no deben estar correlacionados con el factor no observado de la brecha salarial observada en el 2000. Noten, sin embargo, que ello debe valer solo para todas aquellas poblaciones donde hubo efectivamente flujo de migrantes. En palabras más sencillas, lo que establecen ambas condiciones es que la única manera por la cual tener distintas proporciones de inmigrantes de diferentes nacionalidades en 1980 pudo afectar la brecha salarial en los 2000 en una ciudad con características  $X_l$  es porque esa distribución particular atrajo más inmigrantes de los flujos de inmigración observados entre 1980 y el 2000.

**c. Usando el contexto y los datos proporcionados.**

- i. **Estime  $\tau$  por MCO y por MC2E. Incluya controles, y utilice como pesos para las regresiones la población de 1990 ¿por qué es importante incluir los pesos? Muestre el estadístico F para la regresión MC2E. ¿Por qué es importante mostrarlo?**

ii. VARIABLES	(MCO) resgap2	(MCO) resgap4	(MC2E) resgap2	(MC2E) resgap4
relshs	-0.0297*** (0.00583)		-0.0368*** (0.00700)	
relscoll		-0.0576*** (0.00810)		-0.0783*** (0.0127)
ires80	-0.160 (0.149)	-0.224* (0.127)	-0.170 (0.143)	-0.249** (0.122)
nres80	0.135 (0.173)	0.309** (0.144)	0.142 (0.166)	0.339** (0.140)
logsize80	-0.0946*** (0.0317)	-0.0403 (0.0303)	-0.105*** (0.0307)	-0.0582* (0.0329)
logsize90	0.101*** (0.0310)	0.0307 (0.0309)	0.115*** (0.0312)	0.0537 (0.0356)
coll80	0.0991 (0.469)	-0.0551 (0.427)	0.123 (0.462)	-0.00173 (0.422)
coll90	-0.00660 (0.366)	-0.0216 (0.308)	-0.0154 (0.357)	-0.0459 (0.304)
mfg80	-0.225 (0.299)	-0.373 (0.252)	-0.262 (0.287)	-0.377 (0.245)
mfg90	0.193 (0.369)	0.499 (0.302)	0.231 (0.357)	0.483* (0.284)
Constant	-0.128** (0.0596)	-0.0608 (0.0526)	-0.162** (0.0670)	-0.141* (0.0752)
Observations	124	124	124	124
R-squared	0.210	0.386	0.201	0.354

F	46.62	33.22
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

Es importante realizar la regresión por pesos para evitar otorgar un efecto de la migración general sin tener en cuenta los distintos efectos que la cantidad de población de una ciudad puede tener sobre las diferencias en salarios. Entre más población tenga una ciudad, hay más mano de obra, lo que implica que la mano de obra debería ser más barata. Esto es un ejemplo de la heterogeneidad que se presenta al interior de las ciudades por causa de la cantidad de población que esta tiene. Así, es necesario tener esto en cuenta para poder establecer el efecto del choque migratorio en la ciudad.

En las regresiones de MC2E es importante mostrar el estadístico F para poder evidenciar la validez del instrumento. De esa forma, se convence al lector que el instrumento utilizado es el adecuado, y que los resultados obtenidos permiten realizar inferencias sobre la variable a estudiar. En este caso, se observa que el F para ambas regresiones es aceptable, mayor que 10, por lo que existe evidencia para creer que el instrumento puede funcionar para inferir resultados.

**iii. Suponga que se cumplen los supuestos de identificación. Interprete los resultados.**

Suponiendo que se cumplen los supuestos de identificación se cumplen, se entiende que el efecto encontrado por el estimador MC2E permite capturar el efecto causal de la política. En ese caso, se entiende que, en promedio, ante cambios de 1% en el instrumento tipo Barik, esto es incrementos de 1% en el tamaño de mercado que acaparan los inmigrantes del país c en la ciudad l, el residual de la brecha del salario entre inmigrantes con estudios de secundaria y nativos disminuye en 3,7%. En cuanto al residual de la brecha de salario entre población inmigrante con estudios universitarios y nativos, cambios de un 1% en el tamaño de mercado que acaparan los inmigrantes del país c en la ciudad, en promedio, disminuye la brecha la brecha en 7,8%.

**iv. Argumente si los supuestos de identificación son válidos en el contexto del programa.**

El supuesto de relevancia se cumple. La tabla de regresión del inciso anterior muestra que efectivamente el instrumento es relevante para al menos una ciudad k condicional en las características. El supuesto de exogeneidad no debería cumplirse fácilmente, debido a que las ciudades pueden tener características previas que las hagan más atractivas para los migrantes.

Una ventaja importante de los instrumentos tipo Bartik es que se puede testear el supuesto de exogeneidad. Para lograrlo, se puede evaluar si los “shifts” (o aquellos más importantes) se correlacionan con otros factores ´ que pueden explicar diferencias en la brecha salarial en el tiempo. Uno de los posibles ejercicios es el siguiente:

**d. Utilizando los datos provistos:**

**i. Complete la tabla**

VARIABLES	(1) Fracción de inm. de México 1980	(2) Fracción de inm. de Filipinas 1980	(3) Fracción de inm. de El Salvador 1980	(4) Fracción de inm. de China 1980	(5) Fracción de inm. de Cuba 1980	(6) Fracción de inm. de Europa del Oeste + otros 1980	(7) Bartik para secundaria	(8) Bartik para universidad
Log población 1980	0.0541 (0.0474)	0.0541 (0.0474)	0.0541 (0.0474)	0.0541 (0.0474)	0.0541 (0.0474)	0.0541 (0.0474)	592,138*** (205,888)	234,314*** (84,880)
Fracción univ. 1980	-0.545 (0.344)	-0.545 (0.344)	-0.545 (0.344)	-0.545 (0.344)	-0.545 (0.344)	-0.545 (0.344)	-212,512 (2.430e+06)	1.567e+06 (1.068e+06)
Salario res. nativo 1980	0.601 (0.397)	0.601 (0.397)	0.601 (0.397)	0.601 (0.397)	0.601 (0.397)	0.601 (0.397)	2.674e+06 (1.954e+06)	406,447 (662,140)
Salario res. inm. 1980	-0.652 (0.538)	-0.652 (0.538)	-0.652 (0.538)	-0.652 (0.538)	-0.652 (0.538)	-0.652 (0.538)	-3.850e+06 (2.564e+06)	-609,243 (732,597)
Fracción ind. manuf. 1980	0.0589 (0.247)	0.0589 (0.247)	0.0589 (0.247)	0.0589 (0.247)	0.0589 (0.247)	0.0589 (0.247)	-55,774 (1.238e+06)	-95,388 (431,079)
Constant	-0.207 (0.322)	-0.207 (0.322)	-0.207 (0.322)	-0.207 (0.322)	-0.207 (0.322)	-0.207 (0.322)	-3.229e+06** (1.435e+06)	-1.465e+06*** (526,931)
Observations	124	124	124	124	124	124	124	124
R-squared	0.150	0.150	0.150	0.150	0.150	0.150	0.371	0.430

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**i. Interprete los resultados a la luz de los supuestos de identificación. ¿Se refutan o se apoya el cumplimiento de los supuestos?**

La tabla muestra que para migrantes de diferentes países, las ciudades de la muestra parecen no tener condiciones previas que hayan influido en la cantidad de migrantes que recibían. Esto da indicios de una correcta aleatorización, y se apoyan los supuestos mencionados en el literal anterior.

Suponga ahora ustedes quieren replicar este ejercicio para Colombia en aras de entender el impacto que ha tenido la migración venezolana en las brechas salariales. No obstante, ustedes no cuentan con una información tan detallada como la que existe en Estados Unidos. Tan solo cuentan con la siguiente información:

- Fracción de inmigrantes venezolanos en el 2000 en cada municipio. Noten que uno menos esa cantidad es la fracción de inmigrantes de otros países.
- Cantidad de inmigrantes que ingresaron entre el 2000 y el 2020 al país, discriminados únicamente por el hecho de si son venezolanos o no lo son.
- Solo tienen información para la población con título universitario.
- $r_{lj}$  y  $w_{lj}$  como arriba. (Pueden prescindir del  $j$  si quieren)

**e. Bajo ese contexto:**

**i. Plantee un estimador tipo Bartik análogo al de Card**

El estimador tipo Bartik propuesto por Card hace la razón entre la cantidad de migrantes en una ciudad  $l$  respecto a la cantidad total de migrantes del país  $c$  en Estados Unidos. En el caso planteado no se cuenta con información de la cantidad total de migrantes a nivel nacional, ni con información de la población por municipio para el año 2000, motivo por el cual no es posible realizar un instrumento idéntico al planteado por Card. No obstante, el instrumento planteado toma la idea de usar una referencia de migración total para hacer la razón. Debido a que la proporción total de migrantes se puede representar por la ecuación  $mig_{tot} = mig_{ven} + mig_{otro}$ , se toma esta medida como referencia para el calculo del instrumento. Así las cosas, el instrumento planteado es el siguiente:

$$B_{lj} = z_{mv,2000}g_{mv} + z_{mo,2000}g_{mo}$$

$z_{lc,2000}$  representa el share de migración para Venezuela y para otros países para el 2000.  $g_{lc}$  es la cantidad de migrantes venezolanos y de otros países que ingresaron a Colombia entre el 2000 y el 2020.

**ii. Demuestre que en este escenario, usar el Bartik del inciso i es equivalente usar como IV la fracción de inmigrantes venezolanos en el 2000**

En este contexto, debido a que la cantidad de migrantes de otros países son el resultado de la resta 1 menos la fracción de migrantes venezolanos, el instrumento anterior se vería de la siguiente forma.

$$B_{lj} = z_{mv,2000}g_{mv} + (1 - z_{mv,2000})g_{mo}$$

$$B_{lj} = g_{mo} + (g_{mv} - g_{mo})z_{mv,2000}$$

Así, se observa que el instrumento es función completa de la fracción de venezolanos en el 2000.

- iii. **Bajo que condiciones el efecto encontrado es un LATE**
- iv. **Quiénes son los compliers en esta situación**

En este caso, los compliers hacen referencia a los municipios que por causa de las migraciones venezolanas el residuo del salario incremento entre nativos e inmigrantes. Así, se saca a los municipios donde la diferencia hubiera incrementado aún en ausencia de la migración, los municipios donde la diferencia no hubiera incrementado aún en presencia de la migración, y a los municipios donde la diferencia hubiera disminuido en presencia del tratamiento.

### **3. Tercer ejercicio:**

**a.**

La ecuación planteada muestra que un mayor puntaje de matemáticas no necesariamente está relacionado con un mejor salario. Se evidencia que existe un puntaje de matemáticas que maximiza el salario. Luego de esto, se puede observar que valores altos en la prueba presentan salarios iguales a puntajes bajos en la prueba. Por su parte, el puntaje de la prueba de español tiene una relación directa con el salario, puesto que a mayor el puntaje obtenido en la prueba de español se observa un mayor nivel de salario. Así mismo, se observa que ser productor de aceite de oliva incrementa el salario en 20 unidades respecto a otros tipos de productores agrícolas, así como un incremento en el salario de 50 unidades por participar en el programa. Finalmente, se observa que aquellos productores de aceite de oliva que participaron en el programa incrementaron su salario en 30 unidades adicionales respecto a los otros productores que participaron en el programa.

**b.**

Un posible problema derivado de que la variable independiente sea discreta es que no son no paramétricamente identificables. Esto significa que este tipo de variables no permiten reducir el ancho de banda alrededor del tratamiento para acercarlo a cero.