

Nota 7

Experimentos

1 Introducción

Como ya hemos discutido en las notas previas, uno de los principales problemas que evitan que podamos llevar a cabo inferencia causal utilizando nuestro modelo de MCO es el sesgo por variables omitidas. En las notas anteriores analizamos cómo los métodos de diferencias en diferencias y variables instrumentales hacen algunos supuestos para poder identificar el efecto causal de una variable (X_1) sobre la variable dependiente de nuestro modelo (Y). En la nota de variables instrumentales describimos que la clave para poder identificar el efecto causal de X_1 es identificar un instrumento relevante que no tenga el problema de estar correlacionado con el error del modelo (U_i). El error del modelo incluye todas las variables que no son incluidas en nuestra especificación (tanto observables como no observables). El problema de sesgo por variables omitidas se basa en que dicho error incluye alguna variable (observada o no observada; medible o no medible) que genera sesgo.

Por otro lado, el método de diferencias en diferencias consiste en comparar dos grupos que presumiblemente son similares. El supuesto de tendencias paralelas especifica que ambos grupos se hubieran comportado de la misma manera si el grupo afectado por alguna política no lo hubiera sido. En otras palabras, se asume que su comportamiento contrafactual hubiera sido igual que el grupo de comparación si no hubiera sido afectado por la política. Uno de los principales problemas o limitaciones de este método es que normalmente el grupo afectado por la política es seleccionado por alguna razón. Dicha razón forma parte del error del modelo, lo cual genera un sesgo por variables omitidas. Dicho de otra manera, el supuesto de tendencia paralelas no es válido.

El método experimental consiste en formar ambos grupos de comparación eligiendo los participantes de cada grupo de manera aleatoria. Se forma un grupo de tratamiento (T) que será intervenido de alguna manera y un grupo de control (C) que no es intervenido y forma parte del grupo de comparación. La clave de este método es que dado que ambos

grupos son elegidos de manera aleatoria, por construcción no debe haber correlaciones con ninguna variable no observable que forme parte del error.

2 Resultados potenciales

Supongamos que estamos interesado en medir el efecto de distribuir libros durante el verano en el aprendizaje. Por lo tanto, se plantea un experimento que consiste en distribuir libros para que los niños lean durante el verano. Sea Y_i^T el resultado promedio de los niños de la escuela i si reciben libros (i.e. forman parte del grupo de tratamiento) y Y_i^C el resultado promedio de los niños de la escuela i si no reciben libros (i.e. forma parte del grupo de control). A estos valores se les conoce como *resultados potenciales*. A nosotros nos interesa el valor de $Y_i^T - Y_i^C$, sin embargo, solo es posible observar para cada escuela ya sea Y_i^T o Y_i^C . Sea Y_i el promedio observado de los niños de la escuela i .

Lo que si podemos tratar de estimar, bajo algunos supuestos, es $E(Y_i^T - Y_i^C)$, que corresponde al **efecto promedio de tratamiento (Average Treatment Effect, ATE)**. Dado que solo observamos uno de los dos resultados potenciales supongamos que Y_i es el resultado observado y T_i es una dummy que indica si i recibió tratamiento. En este caso tendremos que:

$$Y_i = Y_i^T T_i + Y_i^C (1 - T_i) \quad (1)$$

Para estimar el ATE podemos utilizar los datos observados y estimar:

$$D = E(Y_i|T) - E(Y_i|C) = E(Y_i^T|T) - E(Y_i^C|C) \quad (2)$$

En este caso D corresponde a la diferencia del promedio de escuelas que recibieron el tratamiento y el promedio de escuelas de control. Para que D sea igual al ATE, en la ecuación [2], sumemos y restemos $E(Y_i^C|T)$. Este valor es no observado y corresponde al promedio que las escuelas de tratamiento hubieran tenido si no hubieran recibido el tratamiento.

$$D = [E(Y_i^T|T) - E(Y_i^C|T)] + [E(Y_i^C|T) - E(Y_i^C|C)] \quad (3)$$

El primer término de [3] ($E(Y_i^T|T) - E(Y_i^C|T)$) corresponde al efecto estimado del tratamiento en las escuelas que fueron seleccionadas para el tratamiento (Treatment on the Treated, TOT). El segundo término ($E(Y_i^C|T) - E(Y_i^C|C)$) corresponde al *sesgo por selección*, ya que es la diferencia en el promedio de escuelas de tratamiento y control si ninguna de las escuelas hubiera recibido el tratamiento. Por ejemplo, supongamos que tenemos grupos de escuelas que decidieron distribuir libros y otras que no lo hicieron. Si comparamos los resultados de estos tipos de escuelas, probablemente el sesgo por selección exista. Es decir, si las escuelas que distribuyen libros lo hacen porque los padres de familia están preocupados por las actividades de verano y quieren incentivar a sus hijos a leer, aunque las escuelas no distribuyeran libros, muy probablemente $E(Y_i^C|T) > E(Y_i^C|C)$ ya que los padres del grupo de tratamiento puede ser que encuentren otra forma de incentivar a sus hijos a aprender. Podríamos argumentar el signo contrario de sesgo por selección si se llevo a cabo un programa de gobierno para distribuir libros a los niños mas desaventajados.

Si logramos: (i) tener grupos comparables y (ii) asumir que el hecho de que alguna escuela reciba libros no afecte los resultados de alguna escuela que no los reciba,¹ podremos asumir que el sesgo por selección desaparece ($E(Y_i^C|T) - E(Y_i^C|C) = 0$) y obtenemos:

$$D = E(Y_i^T - Y_i^C|T) = E(Y_i^T - Y_i^C) = ATE \quad (4)$$

Para obtener un estimador del ATE bajo los supuestos establecidos basta con estimar utilizando MCO:

$$Y_i = \tau_0 + \tau_1 T_i + U_i \quad (5)$$

Cabe señalar que en este caso ATE es una estimación del efecto agregado del tratamiento. Dicho de otra manera, no resulta de una derivada parcial (caeteris paribus), sino de una derivada total. Esto se debe a que es muy posible que el tratamiento, además de tener un impacto directo sobre la variable dependiente, puede tener efectos indirectos. En nuestro ejemplo de lectura, puede ser que los niños que reciben libros para leer durante el verano pasen menos tiempo en la calle o vean menos televisión. Esto puede llegar a tener efectos sobre sus resultados, pero será imposible determinar que parte del efecto que identificamos se debe a los efectos directos o indirectos a menos que controlemos por dichas variables.

¹Veremos las implicaciones de violaciones a este supuesto mas adelante

3 Asignación aleatoria

La forma mas sencilla de formar grupos comparables consiste en determinar el status de tratamiento o control de forma aleatoria. Siguiendo este procedimiento, es posible asegurarse que el estatus de tratamiento será independiente de cualquier variable no observable y podemos obtener un estimador insesgado de τ_1 .

En la práctica algunos de los métodos de asignación aleatoria son:

1. Loterías por sobre-demanda. En situaciones en las cuales la demanda para algún programa es demasiado alta y no existen recursos suficientes para servir a toda la población, se considera que distribuir los recursos por lotería puede ser uno de los métodos éticamente más justo. Ejemplos: vouchers en Colombia, préstamos en Sudáfrica.
2. Expansión en fases. Algunos proyectos se expanden en distintas fases a lo largo del tiempo. Para determinar el orden de expansión puede considerarse justo determinarlo de forma aleatoria. Ejemplos: programa de desparasitación en Kenya. Sin embargo, este método puede tener problemas si los individuos anticipan la expansión y reaccionan y no pueden medirse efectos de largo plazo.
3. Asignación aleatoria dentro de grupos. En algunos casos se considera éticamente injusto que algunos grupos reciban beneficios de un programa y otros no. Algunas alternativas consisten en otorgar el beneficio de forma aleatoria a subgrupos dentro de cada grupo. Ejemplo: balsakis en India. Un problema de este método es el supuesto establecido en la sección anterior donde el estatus de otros individuos no te afecte.
4. Fomento al tratamiento. Por último, si negar los recursos puede determinarse no ético, una alternativa consiste en dar acceso general, pero a un grupo de individuos ofrecerles incentivos a participar en el tratamiento. Ejemplos: muestras de fertilizantes en Kenya, en E.U. enviar materiales gratis para estudiar para un examen.

Una prueba que se hace para demostrar que la asignación aleatoria funcionó y que los grupos resultantes son comparables consiste en utilizar información previa al experimento (información basal) y comparar ambos grupos. Generalmente los experimentos incluyen

una tabla en la cual se comparan medias de variables contenidas en la base de datos. En teoría, no debería de poder rechazarse la hipótesis que las medias de cada variable para ambos grupos son iguales.

En la mayoría de los casos alguna de las variables resulta en diferencias significativas, solo por definición probabilística. En estos casos, un ejercicio común consiste en estimar el modelo [5] con y sin estas variables como control y comparar la estimación de τ_1 en ambos casos.

Un método popular para llevar a cabo la asignación aleatoria consiste en agrupar a los individuos por características similares formando grupos (*matched-pairs*). Esto se lleva a cabo utilizando la información basal o datos administrativos preexistentes. Una vez hecho esto se determina de manera aleatoria quien dentro del grupo o *pair* recibe el tratamiento. Este método además de garantizar tener grupos comparables, te permite disminuir la varianza de tu estimador.

4 Problemas de implementación

4.1 Estratificación

En algunos casos la probabilidad de recibir el tratamiento puede depender del cluster en el que te encuentres (ejemplo de vouchers en Colombia). En este caso, el efecto del tratamiento será válido únicamente dentro del grupo. Para estimar el ATE, se tendría que tomar una media ponderada, donde los ponderadores vendrían determinados por variables que determinan las diferencias en probabilidad de recibir tratamiento en los distintos grupos. Otra alternativa es estimar un MCO controlando por estas características. Sin embargo, esta metodología resultará en errores estándar mayores al ejemplo de *matched pairs* debido a que podrían existir grupos que solo tienen individuos control o tratamiento (lo cual agrega ruido y varianza a la estimación).

4.2 Participación parcial

En algunos casos decides de manera aleatoria los individuos que recibirán el tratamiento. En nuestro ejemplo distribuir los libros. Sin embargo, en diversos casos el tratamiento suele ser distinta de la *intención de tratamiento*. En nuestro ejemplo, si la pregunta relevante es ver como leer durante el verano mejora tus resultados, el hecho de leer es distinto de recibir libros, que es lo que impulsa la política. En este caso, supongamos que si el tratamiento (T_i) es leer durante el verano, necesitaremos crear una nueva variable que sea recibir libros (dummy Z_i si la escuela i fue aleatoriamente elegida para recibir libros). En este caso, dos cosas distintas serán el modelo [5] (utilizando T_i) y la siguiente estimación:

$$Y_i = \lambda_0 + \lambda_1 Z_i + V_i \quad (6)$$

Dado que la distribución aleatoria consiste en la distribución de libros, λ_1 será un estimador insesgado. Esto se conoce como *intención de tratamiento* (Intent to Treat, ITT). Sin embargo, la decisión de leerlos o no (T_i) no es aleatoria, por ende, estimar [5] utilizando MCO nos dará un estimador sesgado de τ_1 . En concreto, decidir leer los libros puede ser una decisión basada en el interés en la lectura, por lo tanto, habrá sesgo por variables omitidas. En algunos casos, el estimador λ_1 puede tener un interés en si mismo. Si por ejemplo, quieres ver la eficiencia de este programa de distribución de libros. Sin embargo, si hay distintas maneras de incentivar la lectura y lo que te interesa realmente es estimar τ_1 tendremos que tomar supuestos específicos para obtener un estimador insesgado.

Para poder obtener un estimador insesgado de τ_1 necesitaremos llevar a cabo los siguientes supuestos:

1. Independencia. $\{Y_i^T, Y_i^C\}$ son independientes de Z_i .
2. Monotonicidad. $T_i(1) \geq T_i(0)$. Es decir, no defiers. (Recuerden la tabla de always-takers, never-takers, defiers y compliers)

Bajo estos supuestos podemos utilizar el método de variables instrumentales para identificar el efecto del tratamiento (T_i) para los compliers. Esto se conoce como el *Efecto Promedio de Tratamiento Local* (o Local Average Treatment Effect, LATE). En clase demostraremos que IV identifica el ATE para una subpoblación (los compliers). El estimador

de LATE resulta de aplicar la metodología de IV. Es decir, estimamos la forma reducida [6] y la primera etapa:

$$T_i = \eta_0 + \eta_1 Z_i + W_i \quad (7)$$

El estimador de LATE será $\tau_1 = \lambda_1/\eta_1$. Debemos recordar que Z_i debe cumplir los supuestos de variables instrumentales: (i) exogeneidad, que dado que es distribuido de manera aleatoria, no debe haber problema. Solo cabe señalar que Z_i no debe influir directamente la variable dependiente Y_i , mas que a través de T_i ; (ii) relevancia, que quiere decir que, en promedio, ser asignado aleatoriamente debe hacer mas probable que tomes el tratamiento.

El supuesto más fuerte en este caso es el de independencia. En muchos casos lo más correcto es tomar el ITT como el único estimador que es posible interpretarse de manera insesgada. Tomemos el ejemplo del programa de desparasitación en Kenya. Sea Z_i el indicador de estar en una escuela que aleatoriamente fue elegida para distribuir las pastillas de desparasitación y sea T_i una dummy que indica si el niño efectivamente tomó la pastilla de desparasitación. En este caso, estimando [6] podemos generar un estimador insesgado del ITT, es decir, el efecto del programa. Pero además podríamos estar tentado a calcular el LATE utilizando el IV para estimar el efecto promedio de la desparasitación sobre los niños compliers. Sin embargo, en este caso el supuesto de independencia no es correcto, ya que los niños en las escuelas de tratamiento que hayan decidido no tomar las pastillas (never-takers) se beneficiaron del tratamiento ya que a sus compañeros (algunos de ellos si las tomaron, los compliers) son en promedio menos propensos a propagar enfermedades de parásitos. En este caso, únicamente es correcto estimar el ITT.

4.3 Externalidades

Hay una gran cantidad de intervenciones que provocan externalidades. Por ejemplo, la intervención que distribuyen pastillas desparasitantes beneficia no solo al grupo de tratamiento, sino también al control, ya que la probabilidad de contagio se reduce. Las externalidades pueden generarse por distribución de información, aprendizaje y reacción de los grupos de control. En estos casos, incluso el estimador ITT puede estar sesgado.

Si se espera que las externalidades surjan, el diseño del experimento puede incorporar este componente para estimar el alcance e importancia de dichas externalidades. Un ejemplo de esto lo llevan a cabo Duflo y Saez, que quieren ver el efecto de distribuir información sobre la selección de planes de retiro. Lo que hacen es hacer la asignación aleatoria en dos pasos: (i) elegir de manera aleatoria algunas instituciones donde se distribuiría la información; (ii) dentro de las instituciones seleccionadas, elegir de manera aleatoria individuos a quienes se distribuiría la información. En este caso, los autores estiman la externalidad producto de diseminación de información comparando los resultados de individuos no seleccionados en instituciones seleccionadas con individuos no seleccionados en instituciones no seleccionadas. El supuesto es que aquellos individuos no seleccionados en instituciones seleccionadas se beneficiarían de la información que reciben de sus compañeros que si fueron seleccionados.

4.4 Pérdida de observaciones

La pérdida de observaciones puede ser problemática en el caso de los experimentos. En el caso en el que las observaciones perdidas son aleatorias, este problema implica un menor poder estadístico. Sin embargo, el principal problema se da cuando la pérdida de observaciones no es aleatoria. En particular, si aquellos que se benefician en menor medida del tratamiento deciden abandonar el experimento, llevar a cabo la estimación sin tomar en cuenta la pérdida de observaciones nos puede llevar a sobreestimar el efecto del experimento.

La pérdida de observaciones es costosa y en algunos casos muy difícil de evitar. Es una práctica común reportar en cada experimento que proporción de la muestra basal se ha perdido en las encuestas subsecuentes. En particular, la encuesta basal es útil para poder determinar si las observaciones perdidas son similares a las que se mantienen en el experimento a través de la comparación de sus características observables antes del experimento.

En los casos en que se reconoce que la pérdida de observaciones no es aleatoria es recomendable analizar que tipo de sesgo puede generar este tipo de selección. Además existen algunas estrategias para tratar de controlar este tipo de selección. Una de ellas consiste en hacer un pareamiento (matching) de individuos del tratamiento y del control

utilizando las características recabadas en la encuesta basal.

5 Críticas

- Efectos de equilibrio general. Los experimentos generalmente son de baja escala, lo cual no permite analizar los efectos de equilibrio general que dichas intervenciones implicarían. Estos efectos son importantes para poder evaluar las implicaciones de bienestar que conllevaría la aplicación de las intervenciones como política.
- Efectos en el comportamiento: *Hawthorne* y *John Henry*. La implementación de un experimento puede conllevar cambios en el comportamiento de los individuos. En particular, los individuos que reciben el tratamiento pueden simpatizar con el experimento, saber que son observados y por ende, esforzarse para que haya efectos positivos del experimento. Por otro lado, aquellos que son parte del grupo de control pueden sobreesforzarse para competir con el grupo de tratamiento. Estos comportamientos se conocen como efectos *Hawthorne* y *John Henry*, respectivamente y potencialmente no se hubieran dado en ausencia del experimento.
- Validez externa. Tres preocupaciones que generalmente surgen con los experimentos son:
 - Si el experimento fue desarrollado con amplio nivel de cuidado será difícil asumir que así será llevado a cabo si se generaliza como política pública.
 - El hecho de que el experimento se haya llevado a cabo con una muestra específica genera preocupación de que el mismo resultado se daría con alguna muestra/población distinta.
 - Qué tanto los resultados se deben a detalles específicos de la intervención. Es decir, qué tanto se puede aprender de intervenciones similares.

6 Experimentos naturales

La misma metodología cubierta en esta Nota puede seguirse en el caso de experimentos naturales. Los experimentos naturales se dan cuando algún evento exógeno forma dos gru-

pos: tratamiento (T), es decir, aquellos individuos afectados por el evento y un grupo de control (C), aquellos no afectados. Si el evento es auténticamente exógeno, ambos grupos deben de ser comparables antes de la ocurrencia del evento. Esto puede verificarse, al igual que en un experimento social, haciendo una comparación de medias con datos recabados antes del evento.

Ejemplos de experimentos naturales incluyen:

- Vietnam Era Draft Lottery. Ser elegible para ser llamado al ejército durante la guerra de Vietnam se determinaba por un número que era asignado de manera aleatoria dependiendo del día de nacimiento de la persona. Esta elegibilidad fue utilizada como una fuente de variabilidad exógena para determinar el efecto de estar enlistado sobre los ingresos vitales de los individuos.
- La epidemia de la influenza española de 1918 se utilizó como una variación exógena para determinar la importancia del desarrollo en-útero. Se demostró que aquellos individuos que con alta probabilidad se encontraban durante su desarrollo en-útero durante el lapso de la pandemia tuvieron efectos negativos sobre años de educación, discapacidades físicas, ingreso y estatus socio-económico.
- El mes sagrado del Ramadan se utiliza para determinar la importancia de la salud fetal y la alimentación sobre el desarrollo de los individuos. Se comparan individuos de madres árabes que se desarrollaron en-útero durante estos meses contra individuos que se desarrollaron en otros meses y se encuentran efectos negativos sobre peso al nacer, mortalidad pre-natal y discapacidades en adultos.
- Ubicación de salones. Se comparan resultados de lectura de salones que por su ubicación eran afectados por el ruido del transporte público, respecto a salones no afectados. Se encuentra que la media de lectura en salones expuestos al ruido tenían un rezago equivalente a 3-4 meses de aprendizaje. En este caso, se argumenta que la localización de los salones es una fuente de variación exógena.

7 Tamaño de la muestra y poder estadístico

Llevar a cabo experimentos es costoso. Por esta razón, muchos experimentos vienen acompañados de cálculos de cuál tiene que ser el tamaño de una muestra para lograr identificar un efecto de manera significativa. Este ejercicio se conoce como *cálculo de poder estadístico*.

Para entender en que consiste el cálculo de poder estadístico podemos partir del planteamiento básico de nuestro modelo (ecuación [5]). En este caso, supongamos que el efecto verdadero del tratamiento es τ (por simplicidad asumiremos que $\tau > 0$, pero el caso de $\tau < 0$ es simétrico). El cálculo del poder estadístico querrá determinar de que tamaño tiene que ser una muestra (N) para que exista una alta probabilidad (poder estadístico) de obtener un estimador que nos permita rechazar la hipótesis nula: $H_0 : \tau = 0$ en favor de la alternativa $H_1 : \tau \neq 0$. Para empezar, para poder rechazar la hipótesis nula, tendremos que establecer un nivel de significancia α (o error tipo I). El error tipo I es la probabilidad de rechazar la hipótesis nula si dicha hipótesis es verdadera:

$$\alpha = Pr\left(\frac{\hat{\tau}}{se(\hat{\tau})} > t_\alpha | \tau = 0\right) \quad (8)$$

Nuestro segundo componente es el *poder estadístico*. El *poder estadístico* (κ) es la probabilidad de obtener un estadístico-t mayor a t_α (para tener un estadístico significativo) dado que el valor verdadero de $\tau = \tau_0$ (como se mencionó previamente asumiremos que $\tau_0 > 0$, pero el caso de $\tau_0 < 0$ es simétrico):

$$\begin{aligned} \kappa &= Pr\left(\frac{\hat{\tau}}{se(\hat{\tau})} > t_\alpha | \tau = \tau_0\right) \\ &= Pr\left(\frac{\hat{\tau} - \tau_0 + \tau_0}{se(\hat{\tau})} > t_\alpha | \tau = \tau_0\right) \\ &= Pr\left(\frac{\hat{\tau} - \tau_0}{se(\hat{\tau})} > t_\alpha - \frac{\tau_0}{se(\hat{\tau})} | \tau = \tau_0\right) \\ &= 1 - \Phi\left(t_\alpha - \frac{\tau_0}{se(\hat{\tau})}\right) \end{aligned} \quad (9)$$

Por lo tanto, para obtener un poder estadístico mayor o igual a κ :

$$1 - \kappa \leq \Phi\left(t_\alpha - \frac{\tau_0}{se(\hat{\tau})}\right) \quad (10)$$

$$t_\alpha - \frac{\tau_0}{se(\hat{\tau})} \leq -t_{1-\kappa}$$

Para simplificar, asumamos errores homocedásticos en la estimación de [5] con un tamaño de muestra N y N_T individuos recibiendo el tratamiento, donde $N_T = N * P$ (es decir $100 * P\%$ de los individuos forman parte del tratamiento). En este caso el error estándar del estimador será:

$$se(\hat{\tau}) = \sqrt{\frac{1}{P * (1 - P)} * \frac{\sigma^2}{N}} \quad (11)$$

Por lo tanto, para los valores α , κ , P y N podremos obtener un estimador significativo si:

$$\tau_0 \geq (t_\alpha + t_{1-\kappa}) * se(\hat{\tau}) \quad (12)$$

$$EMD(\tau_0) = (t_\alpha + t_{1-\kappa}) * se(\hat{\tau})$$

donde EMD es el efecto mínimo detectable, es decir, el valor mínimo de τ_0 para el cual podremos obtener un estimador significativo.

- En cuanto al tamaño de muestra N , cabe resaltar que conforme mayor es N , menor es $se(\hat{\tau})$, por lo tanto, el EMD es menor.
- En cuanto mayor es el poder estadístico (κ), mayor es $t_{1-\kappa}$ y por lo tanto, el EMD es mayor.
- Al igual, conforme menor es el error tipo I (α), mayor será t_α y el EMD es mayor.

En casos en los cuales la asignación se haga por grupos (e.g. escuelas), será importante permitir correlación en los errores de individuos en una misma escuela. Para esto se tendrán que asumir errores tipo cluster o el modelo de efectos fijos. En este caso, entre mayor sea la proporción de la varianza explicada por la correlación dentro de un grupo (cluster) (o el intra-class correlation) mayor es el EMD .