

DAVID SALAZAR

EVALUACIÓN DE IMPACTO

LONELINESS

Contents

Introducción 7

Cuasi-experimentos (Experimentos Naturales) 11

Hopefully, a warranted effort.

Introducción

Evaluación de Impacto

El impacto o efecto causal de un programa es simplemente la diferencia entre lo que ocurrió con el programa y lo que habría ocurrido sin el programa. Este último escenario se denomina el contrafactual. Las metodologías de evaluación de impacto van a tratar de aproximar ese contrafactual, a partir de datos observados, en distintos contextos y bajo distintos supuestos.

Si el grupo de control no es verdaderamente comparable con el grupo de tratamiento, se puede subestimar (o sobreestimar) el impacto del programa. Es decir, podemos encontrar un impacto del programa más pequeño (o más grande) de lo que en realidad es. Eso implica que el estimador del impacto estaría sesgado. En estos casos, nuestra aproximación al contrafactual no es correcta.

Potential Outcome Approach

$$D_i = \begin{cases} 1 & \text{si el individuo recibe el tratamiento} \\ 0 & \text{de lo contrario} \end{cases}$$

Luego, los potential outcome del individuo i están dados por: $Y(D_i)$. Por lo tanto, el efecto del tratamiento (evaluación de impacto) para el individuo i está dado por:

$$\tau_i = Y(1) - Y(0)$$

Debido al problema fundamental de la inferencia causal, no podemos observar ambos resultados para un mismo individuo. Sin embargo, podemos comparar. Encontrar una comparación directa entre individuos es difícil; más fácil comparar grupos y promediar el efecto.

$$\tau_{ATT} = E[Y_i(1) - Y_i(0) | D_i = 1]$$

Sin embargo, $E[Y_i(0)|D_i = 1]$ no es observado. Luego, para hacer la comparación, es necesario suponer que el contrafactual del grupo de tratamiento es el siguiente:

$$E[Y_i(0)|D_i = 1] = E[Y_i(0)|D_i = 0]$$

Es decir que, en ausencia del programa, ambos grupos habrían sido iguales. Supuesto equivalente a independencia condicional del tratamiento. Para extrapolar el efecto del tratamiento sobre los tratados sobre cualquier unidad (ATE), tendríamos que hacer el supuesto adicional de que, en promedio, el efecto del tratamiento es el mismo entre tratados y no tratados. Cuando este supuesto no se cumple, hablamos de sesgo de selección: los grupos son sistemáticamente distintos, ya que el grupo de control está seleccionado en características distintas a las del grupo de control; características que ocasionaron que estas no participarán en el programa.

Experimentos Aleatorios

Los experimentos aleatorios son la mejor forma de garantizar que no hay sesgo de selección. Si la asignación aleatoria se hizo correctamente, los dos grupos deberían ser idénticos en promedio antes del inicio de la intervención. De esta manera, se cumple por construcción la condición de identificación. Si la asignación no es aleatoria y los individuos escogen ellos mismos dónde ir, es posible que se segreguen. Si las razones por las cuales se segregan están correlacionadas con la variable de resultado entonces esto es un problema para la estimación del impacto.

Validez

Validez Interna: dada la estrategia de identificación, ¿qué tan creíble es que el estimador del efecto causal en la población y el contexto del estudio?

Validez externa: ¿qué tan generalizables son los resultados a otras poblaciones o contextos?

Non Compliance

Si los individuos no siguen el protocolo, y terminan autoseleccionándose en cualquiera de los grupos, tendremos sesgo de selección. Podemos usar asignación aleatoria, aunque no real, como instrumento para estimar.

Balanceo

Todos los obstáculos en la implementación pueden llevar a que el grupo de control y el de tratamiento difieran sistemáticamente. Una forma de evaluar si esto ocurre es mirar si los dos grupos están ?balanceados? en términos de características observables. Si la aleatorización fue correcta, no deberían existir diferencias significativas en estas características. Por lo tanto, se pueden comparar medias en línea de base.

Modelo de Diferencias

Si los supuestos de aleatorización se cumplen, el ATT se puede estimar simplemente como una diferencia de medias. En regresión lineal:

$$Y_i = \beta_0 + \tau_{ATT}D_i + u_i$$

Modelo que puede ser estimado consistentemente mediante MCO, siempre y cuando estemos dispuestos a hacer el supuesto $E[u_i|D_i] = E[u_i]$

Aunque no hace falta incluir otras variables explicativas, podemos incluir otras que influyan en la variable de interés, disminuir la varianza del error, y ganar precisión en la estimación. Problema? Controles pueden ser endógenos.

De otro lado, podemos incluir controles, tal que, condicional en ellos, no haya sesgo de selección.

Igualmente, podríamos añadir variables al modelo para obtener efectos heterogéneos al modelo.

Cuasi-experimentos (Experimentos Naturales)

Los experimentos aleatorios no siempre son posibles por razones, éticas, prácticas o financieras. En muchos casos, sin embargo, eventos específicos pueden pensarse como si indujeran asignaciones aleatorias que asemejan experimentos aleatorios. En el sentido de que la razón del cambio está dada por razones que no tienen efecto en la variable de interés; luego, podemos tratar la variación causada por el cuasi-experimento como una variación exógena. Por ejemplo: En Carolina del Sur, la estructura fiscal hizo que se terminara el presupuesto que se utiliza para cubrir costos de abortos para pacientes de Medicaid antes de finalizar el año.

Si el EN afecta diferencialmente a las unidades, pero las diferencias son sistemáticas en el tiempo, puede usarse el método de Diferencias en Diferencias que veremos a continuación.

Diferencias en Diferencias

El Modelo de Diff-en-Diff asume que hay un tratamiento diferenciado por unidad, tal que unas unidades están expuestas al tratamiento y otras no; sin embargo, no se supone que los controles son un perfecto contrafactual. Si no que se supone que las diferencias entre el control y los tratados se hubiesen mantenido en ausencia del tratamiento. Luego, el contrafactual es el grupo de tratados antes del tratamiento más el cambio que hubiese tenido en ausencia del tratamiento (el cambio que experimentó el grupo de control. Es decir:

$$\hat{\tau}_{DD} = [E[\tilde{Y}_{post}|D = 1] - E[\tilde{Y}_{pre}|D = 1]] - [E[\tilde{Y}_{post}|D = 0] - E[\tilde{Y}_{pre}|D = 0]]$$

$$\hat{\tau}_{DD} = [E[\tilde{Y}_{post}|D = 1] - E[\tilde{Y}_{pre}|D = 1]] + [E[\tilde{Y}_{post}|D = 0] - E[\tilde{Y}_{pre}|D = 0]]$$

Es decir, el supuesto fundamental es que la diferencia entre los grupos se habría mantenido en ausencia del tratamiento: **Tendencias Paralelas**. Para confirmar la validez de este supuesto, es útil graficar

las variables de interés antes del tratamiento y ver que la diferencia se mantenga constante. Formalmente, un test placebo: DiD cuando sabemos que no ocurrió ningún cambio (no hubo tratamiento); si el supuesto es relevante, el contrafactual debe cancelarse con el valor observado.

Formalmente, el supuesto de tendencias paralelas es:

$$[E[Y(0)_{post} - Y(0)_{pre} | D = 1] = [E[Y(0)_{post} - Y(0)_{pre} | D = 0]]$$

Esto permite la existencia de diferencias entre $D = 1$ y $D = 0$ en no observables siempre y cuando estas sean constantes en el tiempo. En regresión, podemos estimar DiD de la siguiente forma:

$$\Delta Y_i = \alpha + \beta D_i + u_i$$

Errores Estándar

Generalmente, en Panel las observaciones están organizadas por cluster desde el muestreo mismo. Luego, las observaciones no son técnicamente independientes. Si no corregimos esto, estamos sobre-estimando la cantidad de información que brinda la muestra y sub-estimando los errores estándar. Luego, estamos calculando p-values muy bajos. Para corregir por esto, necesitamos corregir por correlación entre los errores dentro de los distintos clusters. Esto es muy importante para la metodología de DiD. Remember: los clusters son las unidades geográficas.

Matching

El método de emparejamiento intenta construir un contrafactual apropiado para cada unidad en el grupo de tratamiento a partir de unidades idénticas o similares en el grupo de control. Esta similitud se basa en características observables. El estimador del ATT por emparejamiento va a ser entonces el promedio de las diferencias en la variable de resultado entre cada unidad tratada y un grupo de control que definiremos para ella con unidades no tratadas ¿comparables?. Luego, el supuesto fundamental es que el sesgo de selección se debe solo variables observables.

Comparar y encontrar "clones" adecuados es complicado a medida que condicionamos en más variables observadas. Bajo el supuesto de que la selección solo depende de variables observables, se pueden emparejar individuos a partir de la probabilidad de estar en el tratamiento condicional en las variables X : $P(D|X)$. Este $P(D|X)$ lo llamamos propensity score matching. Es decir, estamos suponiendo

que es posible construir el contrafactual de un individuo determinado si conocemos su $P(D|X)$. Es decir, $E[Y(0)|D = 1, P(D|X) = \epsilon] = E[Y(0)|D = 0, P(D|X) = \epsilon]$. Supuesto equivalente al de independencia condicional en regresión. Es decir, condicional en el propensity score matching, requerimos que no haya diferencias en los resultados entre control y tratados en ausencia del tratamiento.

Los beneficios de emparejar basados en el propensity score matching y no en todo el espacio de las X : mucho más fácil definir qué es un candidato a clon y qué no.

El sesgo con respecto al contrafactual, luego, será menor a medida de que la $P(D|X)$ sea más similar entre grupo de control y grupo de tratamiento. Luego, solo podremos lograr el matching para los individuos de tratamiento los cuales tengan vecinos en el grupo de control con $P(D|X)$ similares. A esta región de individuos la llamamos región común.

Matching: supone que proporciones en muestra de tratamiento y control son similares a las de la población.

Estimación

1. Estimar $P(D|X)$: Probit o logit. Variables X : teoría económica o ir seleccionándolas basados en la significancia de estas.
2. Definir región de soporte común y desechar observaciones que no califiquen.
3. Definir Algoritmos de Emparejamiento (i.e., qué observaciones no tratadas utilizar como control para cada i).
 - Algoritmos basados en distancia: se define una distancia (relativa a $P_i(D|X)$ máxima dentro de la cuál se considerarán observaciones no tratadas como control. Por ejemplo, Nearest Neighbors, Caliper.
 - Algoritmos de Kernel: definido un Kernel, se ponderan las observaciones de acuerdo al Kernel en la estimación del contrafactual para cada observación. Es decir, dada una observación, calculamos un Kernel que le de un peso a cada observación no tratada de acuerdo a la densidad, y con este peso construiremos la estimación del contrafactual.

Importante: verificar que resultados no sean sensibles a la selección del algoritmo por emparejamiento.
4. Verificar calidad del match:
 - Implementar emparejamiento con X como variable resultado (dado kernels anteriores) y comprobar que no hay efecto

tratamiento. Luego, X 's son similares (grupos balanceados) entre grupo control y tratamiento.

- Al estimar $P(D|X, P(X))$, X debería ser redundante, pues hemos supuesto que $P(D|X)$ capu

Errores estándar

Bootstrapping: muestre con reemplazo de la población, calcule efecto con metodología y guarde resultado. Repita N veces y saque desviación estándar a estos resultados. Este va a ser el error estándar estimado.

Pruebas Placebo

- Ver efecto de tratamiento cuando sabemos que no lo hay: mismos grupos pero antes de tratamiento.
- Ver efecto de tratamiento cuando sabemos que no lo hay: comparar dos grupos que no han sido tratados.

Variables Aleatorias y Promoción aleatoria

En los casos en los que existe un componente aleatorio en la asignación del programa pero aún exista sesgo de selección (por ejemplo si un individuo elegible puede decidir si tomar o no el programa), este componente aleatorio puede usarse como VI. En los casos en los que existe un componente aleatorio en la asignación del programa pero aún exista sesgo de selección (por ejemplo si un individuo elegible puede decidir si tomar o no el programa), este componente aleatorio puede usarse como VI.

Por medio del método de variables instrumentales solo podemos identificar un efecto local. Como el efecto se identifica a partir de los cambios en participación asociados con cambios en la variable instrumental, el efecto que identificamos es para el subconjunto de la población para la cual la VI genera cambios en su decisión de participación. Como el efecto se identifica a partir de los cambios en participación asociados con cambios en la variable instrumental, el efecto que identificamos es para el subconjunto de la población para la cual la VI genera cambios en su decisión de participación.

Cuando un programa ya está implementado, y su asignación no fue aleatoria, no podemos utilizar esta vía. Sin embargo, sí podemos generar una promoción aleatoria del proyecto que podamos usar como variable instrumental del programa. Por ejemplo, se envía una carta a algunos hogares y a otros no. La condición de relevancia

requiere que esta diferencia en la promoción afecte realmente la decisión de participar. La condición de exogeneidad, aparte de la aleatorización, requiere que la promoción no genere efectos sobre la variable de resultado. Debido a LATE, el efecto que hallamos es solo válido para las personas para las cuales la variable instrumental determina qué van a hacer: los que solo participan si son elegibles y les llega la promoción aleatoria.

Regresión Discontinua

Con experimentos naturales, explotamos una variación debido a variables que no están relacionadas con la variable de interés. Luego, consideramos esta variación exógena. Un ejemplo de esto son los umbrales de corte de distintos programas. Quién decide quién participa y quién no, dentro de los individuos que por poco sí y por poco no, es bastante arbitrario. Luego, esta asignación la podemos tratar como exógena (como si fuese producto de asignación aleatoria): los que sí les dieron tratamiento, los que no, no. Luego, debido a que la asignación fue "aleatoria", el grupo de control será un buen contrafactual.

Para explotar este diseño, necesitamos una forcing variable (Z_i) que determina quién es tratado y quién no. Para cumplir que la asignación alrededor del tratamiento si fue como aleatoria, necesitamos los siguientes supuestos:

1. Probabilidad discontinua de sufrir el tratamiento. Es decir, en el punto de corte, cambia la probabilidad de ser tratado. Para ello: grafique con umbral, y compare aproximaciones a ambos lados; debe encontrar una discontinuidad. Si la probabilidad cambia de cero a 1, la regresión discontinua es nítida; si la probabilidad cambia de forma distinta, es regresión borrosa.
2. No debe haber manipulación de la forcing variable. Ver densidad de la forcing variable y ver que no hay cambios raros en el umbral.
3. Los individuos a ambos lados del umbral, dentro del "vecindario", deben ser comparables. Ver que la única discontinuidad sea en la variable de interés; no en ninguna de las explicativas. No es verificable garantizar que hay continuidad en las variables no observables.

RD design

In the simplest version of this design, units are assigned a score, and a treatment is given to those units whose value of the score exceeds a known cutoff and withheld from units whose value of the score is

below the cutoff. This discontinuous change in the treatment assignment probability can be used to infer the effect of the treatment on an outcome of interest because, under certain assumptions, it makes units with scores barely below the cutoff comparable to units with scores barely above it.

In the RD design, all the units in the study receive a score, and a treatment is assigned to those units whose score is above a known cutoff and withheld from those units whose score is below the cutoff. These three components—score, cutoff, and treatment—define RD designs in generality. When all units in the study comply with the treatment condition they have been assigned, we say that the RD is sharp. In contrast, when some of the units fail to receive the treatment despite having a score above the cutoff and/or some units receive the treatment despite having been assigned to the control condition, we say that the RD design is fuzzy.

That is: we adopt a randomized experiment vocabulary, where the treatment is assigned with the value of a score and a cutoff. There can be compliance or not.