

Técnicas de Evaluación de Políticas Públicas

Manuel Arellano

Temas de Economía Empírica

CEMFI, 2003–2004

1. Análisis descriptivo: Introducción y ejemplos

- Se pueden distinguir dos tipos de investigación empírica:
 - (a) Análisis descriptivo.
 - (b) Inferencia causal.
- Los confines entre una y otra categoría son borrosos. No obstante, hay buenos ejemplos de investigación firmemente asentados en cada una de ellas.
- La distinción entre (1) y (2) no depende de la sofisticación de las técnicas estadísticas empleadas. A veces se asocia el adjetivo “descriptivo” al cálculo de cuadros con medias o correlaciones, mientras que se reservan los adjetivos “económico” o “riguroso” para los cálculos de coeficientes de regresión u otros estadísticos más complejos del mismo estilo.
- Una simple comparación de medias puede ser causal, mientras que complejos análisis estadísticos pueden ser descriptivos¹.
- Quizás los éxitos más grandes de la econometría se encuentran en el análisis descriptivo.

¹ Por ejemplo, los análisis de las distribuciones salariales de Buchinsky (1994) o DiNardo, Fortin y Lemieux (1996) utilizando, respectivamente, regresión cuantílica censurada o distribuciones empíricas no paramétricas.

- Ejemplos recientes incluyen tendencias en las tasas de pobreza, participación laboral, desigualdad y movilidad salarial, y flujos migratorios.
- La descripción útil no es un ejercicio mecánico. Se trata de un valioso ejercicio de investigación a menudo asociado con ideas innovadoras. Las ideas tienen que ver con la elección de los aspectos a describir, la forma de describirlos y su interpretación.
- A continuación desarrollo algunos ejemplos para resaltar la importancia de la elección de los aspectos a medir.

Ejemplo 1: Desigualdad salarial en la década de los 80

- Juhn, Murphy y Pierce (1993) estudian los datos de salarios individuales correspondientes a 27 años de la Encuesta de Población de EE. UU.
- Entre 1963 y 1989 los salarios reales aumentaron un 20%, pero con una distribución desigual. Los del percentil 10 (trabajadores menos cualificados) cayeron un 5% mientras que los del percentil 90 (los más cualificados) aumentaron un 40%.
- Juhn et al. repitieron los cálculos por categorías de educación y experiencia. Observaron que la desigualdad salarial también aumentó dentro de cada uno de estos grupos, especialmente en los años 80.
- Finalmente, observaron que las diferencias salariales entre grupos de educación y experiencia también aumentaron considerablemente.
- Juhn et al. interpretan estos resultados como el reflejo de incrementos de las primas de los componentes observables e inobservables de la cualificación (educación, experiencia y talento).
- Este trabajo y otros en la misma línea, documentando aspectos relevantes de los cambios en la distribución de salarios, han generado abundante investigación sobre las causas últimas de estos cambios.

Ejemplo 2: Precio de la vivienda y cambios en la calidad

- Para medir la inflación de precios de la vivienda se recogen precios por metro cuadrado de viviendas vendidas. Se calculan precios medios para distintos momentos del tiempo y la inflación se mide por las diferencias respecto a los precios medios de un año base.
- Un inconveniente de esta medida es que las casas vendidas en distintos momentos del tiempo pueden no ser comparables por mejoras o disminuciones de su calidad. Esta preocupación es mayor en épocas de crecimiento, en que los cambios en la calidad pueden ser importantes. En este caso puede interpretarse erróneamente como inflación lo que en parte son mejoras de calidad en las viviendas.
- Una respuesta a esta preocupación es utilizar medidas de las características de las viviendas para obtener los cambios en los precios de grupos de viviendas con las mismas características. A continuación, se calcula una inflación media de todos los grupos de viviendas. A las técnicas de este tipo se las denomina *métodos hedónicos*.
- Un problema de los precios hedónicos es que no son convincentes si hay características importantes de las viviendas que no se observan, como suele ocurrir con variables de localización (eg. proximidad a colegios o al transporte público).

- Bover y Velilla (2002) estudian datos de más de 100.000 viviendas nuevas españolas vendidas en las principales ciudades entre 1993 y 1997. Estas viviendas se agrupan en conjuntos o “promociones” que comparten las características más importantes. Además estas promociones están en el mercado un promedio de 18-24 meses y contienen viviendas de diferentes tamaños.
- Bover y Velilla utilizan esta circunstancia para calcular medidas de inflación específicas de cada promoción, que a continuación agregan en índices de inflación por ciudades o zonas de residencia. De esta forma la medida de inflación no se ve afectada por diferencias en las características de las viviendas a lo largo del tiempo, al menos de aquellas que son comunes a las promociones, o que se pueden medir dentro de cada promoción.
- Para un conjunto de 6 ciudades para las que Bover y Velilla disponen de los mejores datos, el índice no ajustado aumenta un 3% en el periodo 1993-1997, mientras que el ajustado disminuye un 2%.

Ejemplo 3: La desigualdad mundial de la renta

- Cuando se analizan los cambios en la desigualdad de la renta dentro de cada país se observa un aumento de la desigualdad (aunque no uniforme) en el periodo 1970-1998.
- Cuando se analiza la desigualdad de renta entre países también se observa un aumento en los últimos 30 años.
- Sala-i-Martin (2002) argumenta que de las dos premisas anteriores no se sigue necesariamente que la desigualdad mundial haya aumentado. El motivo es que las medidas de desigualdad *dentro* de cada país y *entre* países no son comparables. Las primeras se refieren a “personas” mientras que las segundas se refieren a “países”.
- Supongamos que 5/6 de la población mundial viven en países con bajo PIB per cápita y ausencia de crecimiento, y 1/6 viven en países con alto PIB per cápita y altas tasas de crecimiento. El resultado es que la varianza de PIB per capita entre países aumentará en el tiempo.
- Supongamos ahora que un país pobre pero grande (con 1/6 de la población mundial, ie. China) tiene altas tasas de crecimiento, de forma que las rentas de sus habitantes tienden a converger con las de los países ricos. Como este país contribuye una *única observación*, la varianza entre países continúa aumentando.

- Simultáneamente, este país, que previamente tenía una distribución personal de la renta muy comprimida, empieza a experimentar un ensanchamiento de la distribución por lo que aumenta la desigualdad dentro del país.
- Sin embargo, las rentas de 1/6 de la población mundial se están aproximando a las de los ricos, por lo que puede ocurrir que la reducción de la distancia entre estos 1200 millones de personas y los ricos, compense el aumento de la desigualdad dentro del país, dando lugar a una reducción de la desigualdad mundial de las rentas individuales.
- Sala-i-Martin asigna un nivel de renta a cada persona en el mundo y estima la distribución mundial de la renta para el periodo 1970-1998.
- A continuación calcula tasas mundiales de pobreza integrando la distribución por debajo de líneas de pobreza dadas. Concluye que hay entre 300 y 500 millones menos de pobres en el mundo en 1998 que en 1970.
- También calcula distintas medidas de desigualdad de la renta. Todos los índices muestran una reducción de la desigualdad de la renta global entre 1980 y 1998. Las reducciones se deben sobre todo a reducciones de la desigualdad entre países, debidas en especial pero no únicamente a las altas tasas de crecimiento de las rentas de los 1200 millones de habitantes de China.

2. Programas de formación y empleo asistido

- La evaluación de las políticas de formación ha sido objeto de interés creciente con la expansión del gasto en políticas activas de mercado de trabajo en los últimos 20 años.
- Estudiaremos de forma detallada una investigación concreta (Ham y LaLonde, 1996) a fin de extraer conclusiones generales sobre los aspectos siguientes:
 - (a) Variables de evaluación.
 - (b) Utilidad de los datos experimentales.
 - (c) Efectos causales en contextos económicos y sociales.
 - (d) Papel de la modelización econométrica.

2.1. Programa y datos

- El *National Supported Work Program* (NSW) se diseñó en los EE.UU. para proporcionar formación y oportunidades de empleo a personas con dificultades para encontrar empleo, como parte de un diseño experimental que permitiera mostrar el impacto de este tipo de políticas.
- Ham–LaLonde estudian el impacto del NSW sobre las mujeres que se presentan voluntarias.
- Requisitos para solicitar la participación en el NSW:
 - (a) Parado al menos tres meses.
 - (b) Haber recibido transferencias del AFDC (*Aid to Families with Dependent Children*) durante unos 30

meses (el AFDC estaba dirigido a familias carentes de apoyo paterno, bien por muerte del padre, ausencia continuada, incapacidad o desempleo).

- (c) No tener hijos en edad preescolar.
- Entre 1976 y 1977 las solicitantes fueron asignadas aleatoriamente a grupos de tratamiento y control. El experimento tuvo lugar en siete ciudades. En el momento inicial (*baseline*) se recogió información sobre edad, educación, abandono escolar (*dropout*), número de hijos, estado civil, raza, e historial laboral en los dos años previos.
 - Historiales de empleo posteriores a la entrada en el programa: “tratados” y “controles” fueron entrevistadas a intervalos de nueve meses recabando información sobre estatus laboral. De esta forma se consiguieron duraciones de periodos de empleo y desempleo tras la entrada durante más de dos años (26 meses).
 - En los datos usados por Ham–LaLonde hay 275 mujeres en el grupo de tratamiento y 266 en el de control. Todas se presentaron voluntarias en 1976. Características medias: edad 34, 10 años de educación, 70% de *dropouts*, 2 niños, 65% casadas, 85% de raza negra.
 - NSW garantizaba 12 meses de empleo subvencionado (en “contrato de formación”) en empleos en los que la productividad aumentaba gradualmente en el tiempo.

2.2. Variables de evaluación

2.2.1. Tasas de empleo

- Gracias al diseño experimental, una simple comparación entre las tasas de empleo de los grupos de tratamiento y control proporciona una estimación insesgada del efecto del programa sobre la probabilidad de empleo de los tratados (*average treatment effect on the treated*).
- La Figure 1 de Ham–LaLonde nos da la información básica.
- El crecimiento en las tasas de empleo de los controles refleja los criterios de elegibilidad al programa.
- La conclusión de la evaluación experimental es que al menos a corto plazo NSW mejoró sustancialmente las perspectivas de empleo de los participantes del AFDC (una diferencia de 9 puntos porcentuales en las tasas de empleo).
- Credibilidad de las evaluaciones no experimentales. LaLonde (1986) utilizó PSID y CPS para ver si los métodos econométricos podían generar conclusiones similares a las experimentales con NSW. No consiguió replicar los resultados experimentales, aunque posteriormente Deheja and Wahba (1995) sí que los reprodujeron.

2.2.2. Salarios

- La comparación de salarios medios de tratados y controles proporciona una estimación sesgada del efecto del programa sobre los salarios. Esto ocurrirá en la medida en que el tratamiento (la formación) tenga un efecto sobre las tasas de empleo de los tratados.
- Notación:

$$D = \begin{cases} 1 & \text{tratados} \\ 0 & \text{controles} \end{cases}$$
$$\eta = \begin{cases} 1 & \text{título secundaria} \\ 0 & \text{abandono secundaria} \end{cases}$$
$$Y = \begin{cases} 1 & \text{empleado} \\ 0 & \text{no empleado} \end{cases}$$
$$W = \text{salario para los empleados}$$

- Supongamos que el tratamiento aumenta las tasas de empleo de abandonos y graduados

$$\Pr(Y = 1 \mid D = 1, \eta = 0) > \Pr(Y = 1 \mid D = 0, \eta = 0)$$
$$\Pr(Y = 1 \mid D = 1, \eta = 1) > \Pr(Y = 1 \mid D = 0, \eta = 1)$$

- pero el efecto es de menor intensidad para los graduados:

$$\frac{\Pr(Y = 1 \mid D = 1, \eta = 0)}{\Pr(Y = 1 \mid D = 0, \eta = 0)} > \frac{\Pr(Y = 1 \mid D = 1, \eta = 1)}{\Pr(Y = 1 \mid D = 0, \eta = 1)}$$

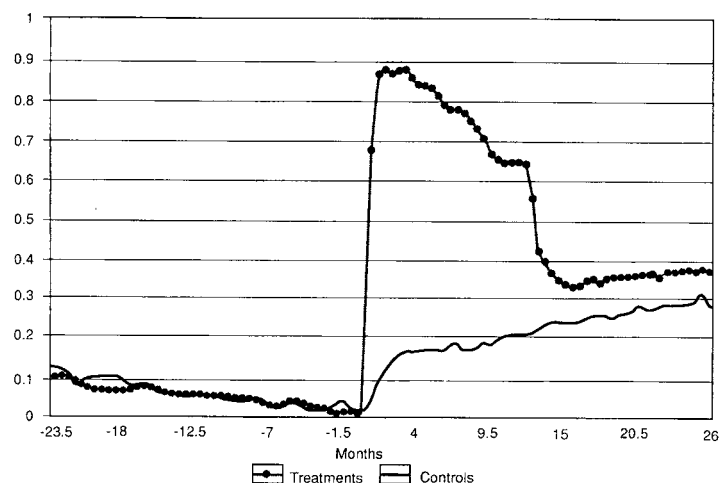


FIGURE 1.—Employment rates of AFDC women in the NSW Demonstration.

experimental evaluation shows that at least in the short run, NSW substantially improved the employment prospects of AFDC participants.

The NSW demonstration achieved these employment gains by helping trainees to hold on to their jobs longer and/or to find jobs faster, thereby increasing the length of their employment spells and/or reducing the length of their unemployment spells. To begin our analysis of these effects of training, we examine the Kaplan-Meier survivor functions for the treatments' and controls' employment and unemployment spells in Table I.⁶ The first two columns of the table indicate that 65 percent of the trainees' employment spells lasted six or more months compared with only 57.3 percent of the controls' spells. When we follow standard practice and compare the experience of treatments and controls in fresh unemployment spells in columns three and five of Table I, we see that 73 percent of the treatments are still in an unemployment spell after a duration of 6 months compared to only 61.3 percent of the controls. Thus training appears to be a mixed blessing since it increases the length of both employment and unemployment spells.

Unfortunately, as previously noted, such a simple analysis of the treatments' and controls' employment histories may be misleading. First, the possibility that the treatments and controls faced different demand conditions is particularly

⁶ In practice many of the employment and unemployment spells are not completed during the sample period (i.e., they are right censored). Therefore, we cannot simply compare their mean durations, especially because the treatments spend on average half the sampling frame in training.

TABLE I
EMPIRICAL SURVIVOR FUNCTIONS
(Proportion Remaining Employed or Unemployed)

Months	Employment		Unemployment		
	Treatments (1)	Controls (2)	Treatments (3)	Controls: All Spells (4)	Controls: Fresh Spells (5)
1/2	0.968 (.013)	0.929 (.018)	0.955 (.013)	0.949 (.011)	0.929 (.023)
1	0.929 (.019)	0.848 (.026)	0.910 (.018)	0.914 (.015)	0.895 (.028)
2	0.839 (.027)	0.761 (.030)	0.864 (.021)	0.843 (.019)	0.791 (.039)
3	0.787 (.031)	0.687 (.033)	0.817 (.024)	0.807 (.021)	0.756 (.039)
4	0.733 (.033)	0.648 (.034)	0.778 (.025)	0.781 (.022)	0.728 (.039)
5	0.670 (.034)	0.603 (.034)	0.746 (.026)	0.756 (.022)	0.672 (.041)
6	0.650 (.035)	0.573 (.035)	0.730 (.027)	0.725 (.023)	0.613 (.043)

Notes: The calculations in Column 4 include spells in progress at the baseline. (In the spells in progress, duration is measured from the baseline.) Those in Column 5 use only unemployment spells that begin after the baseline. The standard error calculations account for "right censoring" of the data.

pertinent in the present case because participants entered NSW as the economy was recovering from the 1974-75 recession. As a result, the controls encountered significantly worse labor market conditions during the portions of their spells that occurred while the treatment group received training.⁷

More importantly, as also observed in Section 2, an experimental design ensures only that the entire sample of treatments and controls are random draws from the same population. It does not ensure, for example, that the subsamples of treatments and controls experiencing employment spells are drawn randomly from the same population. To explore this possibility, we present in Table II the mean characteristics of treatments and controls experiencing different types of spells. As shown by columns one and two, these figures suggest that treatments experiencing employment spells are younger, less skilled, less likely to have ever been married, and have fewer weeks of work in the previous two years. Because these characteristics usually are associated with shorter employment durations, the NSW demonstration may have had an even larger impact on these durations than that suggested by the Kaplan-Meier estimates in Table I.

A similar problem arises in the fresh unemployment spells. As shown by columns three and five of Table II, the treatments and controls experiencing

⁷ As noted in Section 2, we could avoid this problem (but encounter others) by using only the post-training data for treatments and controls. We do not focus on this problem here since standard parametric models should be able to deal with it.

- Se puede comprobar que esto implica que la fracción de abandonos será mayor en el grupo de empleados tratados que en el de empleados controles:

$\Pr(\eta = 0 \mid Y = 1, D = 1) > \Pr(\eta = 0 \mid Y = 1, D = 0)$,
i.e. η no es independiente de D dado $Y = 1$, aunque incondicionalmente $\eta \perp D$.

- La implicación es que la educación media de los empleados tratados será inferior a la de los empleados controles.
- Por ello, la comparación directa de salarios medios entre controles y tratados tenderá a subestimar el efecto del tratamiento sobre los salarios:

$$\Delta_f = E(W \mid Y = 1, D = 1) - E(W \mid Y = 1, D = 0),$$

- mientras que los efectos de interés de D sobre W son los siguientes:

Para abandonos:

$$\Delta_0 = E(W \mid Y = 1, D = 1, \eta = 0) - E(W \mid Y = 1, D = 0, \eta = 0),$$

para graduados:

$$\Delta_1 = E(W \mid Y = 1, D = 1, \eta = 1) - E(W \mid Y = 1, D = 0, \eta = 1)$$

y el efecto global:

$$\Delta_s = \Delta_0 \Pr(\eta = 0) + \Delta_1 \Pr(\eta = 1).$$

- Por lo general, tendremos que $\Delta_f < \Delta_s$.

- Comentarios

- (a) No es posible construir un experimento que permita medir directamente el efecto de la formación sobre los salarios. Esto es, no parece posible que la correlación condicional entre D y η se pueda deshacer experimentalmente.
- (b) El problema se plantea porque η puede ser inobservable.
- (c) A este tipo de efecto no se le denominaría “causal” en el sentido de Rubin y Holland, en la medida en que no hay un experimento conceptual al que se pueda asociar.
- (d) A pesar de ello, los datos experimentales son útiles para medir el efecto de D sobre W . Con datos experimentales sólo tenemos que modelizar el proceso de selección en el empleo utilizando una estrategia no experimental, mientras que con datos no experimentales tendríamos que modelizar tanto la selección en el empleo como la selección en la formación.

2.2.3. Duraciones de empleo y desempleo

- En el efecto de la formación sobre las tasas de empleo se mezclan dos efectos distintos que tienen interés por si mismos:

(a) sobre la duración del empleo

(b) sobre la duración del desempleo

- ¿A través de qué mecanismo aumenta el programa las tasas de empleo? ¿Ayudando a las participantes a encontrar empleo o ayudándoles a mantenerlos?
- Efectos sobre la duración del empleo. Al igual como ocurría con los salarios, la comparación experimental de tasas de salida del empleo puede ser errónea: Sea T_e la duración de un período de empleo. La comparación experimental nos permite calcular

$$\Pr(T_e = t \mid T_e \geq t, D = 1) - \Pr(T_e = t \mid T_e \geq t, D = 0)$$

pero realmente nos interesa

$$\Pr(T_e = t \mid T_e \geq t, D = 1, \eta) - \Pr(T_e = t \mid T_e \geq t, D = 0, \eta).$$

- D está correlacionado con η dado $T_e \geq t$ por varias razones. e.g. Si el tratamiento ayuda especialmente a aquellos con $\eta = 0$ a encontrar empleo, la fracción de $\eta = 0$'s en el grupo $\{T_e \geq t, D = 1\}$ aumentará en relación a la del grupo $\{T_e \geq t, D = 0\}$.

- Problemas similares se plantean con las duraciones del desempleo. Ham y LaLonde lo resuelven con un modelo econométrico de historiales laborales con heterogeneidad inobservable.
- La conclusión es que la formación aumenta significativamente la duración de los períodos de empleo, pero no tiene efecto sobre la duración de los períodos de desempleo.
- El modelo econométrico consigue eliminar las “anomalías empíricas” (efectos experimentales opuestos a lo esperado sobre las duraciones de paro), pero la credibilidad de los resultados no es comparable a la fuerza de los resultados experimentales.

3. Diferencias en diferencias

3.1. Salarios mínimos y empleo (Card y Krueger, 1994)

- En Marzo de 1992 el Estado de New Jersey elevó un 19% el salario mínimo legal (de 4,25 a 5,05 dólares por hora), mientras que el Estado fronterizo de Pennsylvania lo mantuvo constante.
- Card y Krueger (1994) decidieron evaluar el efecto de este cambio sobre el empleo de los trabajadores de salarios bajos. En un modelo competitivo el resultado de aumentar el salario mínimo es reducir el empleo.
- Para ello hicieron una encuesta a unos 400 restaurantes de comida rápida de los dos estados justo antes de la reforma en New Jersey, y una segunda encuesta a los mismos establecimientos 7 – 8 meses después.

- Características de los restaurantes de comida rápida:
 - (a) Se encuentran entre las principales fuentes de empleo de trabajadores de salarios bajos, que es precisamente el grupo que puede verse afectado por un aumento de los salarios mínimos.
 - (b) Cumplen la legislación de salarios mínimos (especialmente las franquicias, que son un % alto del sector).
 - (c) Las características del empleo son bastante homogéneas, lo que facilita la obtención de medidas fiables de empleo y salarios (además no hay propinas).
 - (d) Es fácil construir un diseño muestral de restaurantes en régimen de franquicia (usando las páginas amarillas), que además tienen unas tasas elevadas de respuesta a encuestas telefónicas—excepto McDonalds, que fueron excluidos.
 - (e) Los autores obtuvieron tasas de respuesta del 87% y del 73% (menor en Penn, porque el entrevistador era menos persistente). Casi todos los que respondieron en la primera ola también lo hicieron en la segunda.

- El coeficiente de diferencias de diferencias es

$$\beta = [E(Y_2 | X = 1) - E(Y_1 | X = 1)] - [E(Y_2 | X = 0) - E(Y_1 | X = 0)].$$

donde Y_1 e Y_2 representan el empleo en los periodos pre y post-reforma de la ley, $X = 1$ indica un establecimiento de New Jersey (grupo de tratamiento) y $X = 0$ de Pennsylvania (grupo de control).

- β mide la diferencia entre el cambio medio del empleo en New Jersey y el cambio medio del empleo en Pennsylvania, o equivalentemente el cambio del periodo 1 al 2 en las diferencias medias de empleo entre los dos estados.
- El supuesto clave que permite dar una interpretación causal a β es que el efecto temporal sobre los dos estados es el mismo en ausencia de intervención. No obstante, es posible generalizar la comparación de varias formas, por ejemplo incluyendo otras variables en la comparación.
- Card and Krueger encontraron que el aumento en el salario mínimo aumentó el empleo en algunas de sus comparaciones y que en ningún caso se redujo el empleo. Este artículo promovió un amplio debate profesional y político.

3.2. El contexto de las comparaciones DD

- Si tenemos datos antes y después de un tratamiento, se podrían comparar los resultados de los tratados tras el tratamiento con los resultados de los tratados antes del tratamiento (i.e. los tratados antes del tratamiento utilizados como controles para los tratados después del tratamiento).
- El problema es que esta comparación está contaminada por el efecto sobre los resultados de sucesos distintos al tratamiento que hayan ocurrido entre ambos periodos.
- Supongamos que *solo una fracción* de la población es expuesta al tratamiento. En tal caso, podemos utilizar al grupo de los que nunca reciben tratamiento para identificar la variación temporal en los resultados que *no se debe* a la exposición al tratamiento. Esta es la idea básica del método de diferencias en diferencias.
- Notación para $t = 0, 1$:
 $Y_{0i}(t)$ = resultado de i en t en ausencia de tratamiento.
 $Y_{1i}(t)$ = resultado en caso de tratamiento entre $t - 1$ y t .
 $Y_{1i}(t) - Y_{0i}(t)$ = efecto causal.
 $D_i(t)$ = Indicador de tratamiento entre $t = 0$ y $t = 1$.
 Resultados observados:

$Y_i(0) = Y_{0i}(0)$ (el tratamiento solo se administra en $t = 1$)

$$Y_i(1) = \begin{cases} Y_{0i}(1) & \text{si } D_i(1) = 0 \\ Y_{1i}(1) & \text{si } D_i(1) = 1 \end{cases}$$

Efecto medio sobre los tratados:

$$\alpha_1 = E(Y_{1i}(1) - Y_{0i}(1) \mid D_i(1) = 1).$$

- El *supuesto de identificación fundamental* es que los cambios medios de los dos grupos (tratados y controles) son iguales en ausencia de tratamiento:

$$\begin{aligned} E(Y_{0i}(1) - Y_{0i}(0) \mid D_i(1) = 1) \\ = E(Y_{0i}(1) - Y_{0i}(0) \mid D_i(1) = 0). \end{aligned}$$

Las variables $Y_{0i}(0)$, $Y_{0i}(1)$ y $Y_{0i}(0)$ se observan, pero $Y_{0i}(1)$ es contrafactual.

- Consideremos la siguiente regresión

$$Y = \beta_0 + \beta_1 D + \beta_2 t + \beta_3 D \times t + v$$

donde $D = 1$ para el grupo que recibe tratamiento en $t = 1$, y $D = 0$ para el grupo que nunca recibe tratamiento (por tanto $D_i(1) = D_i \times t_i$).

- Tenemos:

$$\begin{aligned} \bar{Y}(D = 0, t = 0) &= \beta_0 \\ \bar{Y}(D = 1, t = 1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \\ \bar{Y}(D = 0, t = 1) &= \beta_0 + \beta_2 \\ \bar{Y}(D = 1, t = 0) &= \beta_0 + \beta_1, \end{aligned}$$

por tanto

$$\begin{aligned} \beta_3 &= [\bar{Y}(D = 1, t = 1) - \bar{Y}(D = 1, t = 0)] \\ &\quad - [\bar{Y}(D = 0, t = 1) - \bar{Y}(D = 0, t = 0)] \end{aligned}$$

o bien

$$\begin{aligned} \beta_3 &= E[Y_1(1) - Y_0(0) \mid D(1) = 1] \\ &\quad - E[Y_0(1) - Y_0(0) \mid D(1) = 1]. \end{aligned}$$

Utilizando el supuesto de identificación, reemplazamos el segundo término para obtener que $\beta_3 = \alpha_1$:

$$\begin{aligned} \beta_3 &= E[Y_1(1) - Y_0(0) \mid D(1) = 1] \\ &\quad - E[Y_0(1) - Y_0(0) \mid D(1) = 0] \\ &= E[Y_1(1) - Y_0(1) \mid D(1) = 1] \equiv \alpha_1. \end{aligned}$$

Por tanto, el coeficiente de la interacción $D \times t$ en la regresión anterior proporciona el “efecto medio del tratamiento para los tratados”.

3.3. Datos de panel

- Para calcular el parámetro de DD no necesitamos datos de panel, basta con disponer de datos de corte transversal para al menos $t = 0$ y $t = 1$.
- Si tenemos datos de panel (e.g. como Card y Krueger) podemos considerar la regresión en primeras diferencias

$$\Delta Y = \beta_2 + \beta_3 D + \Delta v.$$

Esta regresión resulta conveniente para tener en cuenta la dependencia de las observaciones entre los dos períodos.

3.4. Problemas

- Diferencias de composición en las poblaciones de corte transversal a lo largo del tiempo (especialmente problemático si no se usan datos de panel).

- Dinámicas no paralelas (no cumplimiento del supuesto fundamental).
- Podría ocurrir que el supuesto fundamental se cumpla de forma condicional dadas ciertas variables, pero si alguna de ellas es *inobservable* la identificación desaparece.
- Otro problema es la posible endogeneidad de los cambios legislativos (Besley y Case, 1994).

3.4. Comentarios

- La estimación DD se ha convertido en un método muy popular para obtener efectos causales, especialmente en los EE.UU. donde la estructural federal da lugar a variación en la aprobación de nuevas leyes entre estados.
- Se trata de un método atractivo por su simplicidad y porque a veces resulta convincente.
- Bertrand, Duflo y Mullainathan (2001) consideraron el problema de sesgos en los errores estándar a causa de ignorar la correlación serial. El problema es potencialmente importante porque las estimaciones DD con frecuencia se basan en series largas (un promedio de 16 períodos en el estudio de Bertrand) y las variables de resultado suelen estar autocorrelacionadas.
- El problema tiene fácil solución pero llama la atención el escaso impacto de la econometría formal en los estudios empíricos analizados por Bertrand et al.

4. Métodos de emparejamiento (*matching*)

- Hay muchas situaciones en las que la realización de experimentos no es posible o es inmoral. Un ejemplo clásico es el análisis de los efectos de fumar sobre la mortalidad.
- Los datos experimentales pueden garantizar la condición de independencia

$$(Y_1, Y_0) \perp D$$

pero con datos no experimentales esta condición es en general poco plausible.

4.1. Independencia condicional

- Una condición menos exigente para datos no experimentales es:

$$(Y_1, Y_0) \perp D \mid X$$

- Siendo Y el resultado observado

$$Y = DY_1 + (1 - D)Y_0$$

la independencia condicional garantiza

$$\begin{aligned} E(Y_1 \mid X) &= E(Y_1 \mid D = 1, X) = E(Y \mid D = 1, X) \\ E(Y_0 \mid X) &= E(Y_0 \mid D = 0, X) = E(Y \mid D = 0, X). \end{aligned}$$

- Por tanto, podemos calcular

$$\begin{aligned} \alpha_{ATE} &= E(Y_1 - Y_0) = \int E(Y_1 - Y_0 \mid X) dF(X) \\ &= \int [E(Y \mid D = 1, X) - E(Y \mid D = 0, X)] dF(X) \end{aligned}$$

$$\begin{aligned}\alpha_{SATE} &= E(Y_1 - Y_0 | D = 1) \\ &= \int E(Y_1 - Y_0 | X) dF(X | D = 1) \\ &= \int [E(Y | D = 1, X) - E(Y | D = 0, X)] dF(X | D = 1).\end{aligned}$$

4.2. Regresión múltiple

- El punto anterior es simplemente la noción básica que subyace la regresión múltiple. Por tanto, si por ejemplo especificamos $E(Y | D, X)$ como una regresión lineal en D , X y DX tenemos

$$E(Y | D, X) = \beta D + \gamma X + \delta DX$$

y de aquí

$$E(Y | D = 1, X) - E(Y | D = 0, X) = \beta + \delta X.$$

$$\alpha_{ATE} = \beta + \delta E(X)$$

$$\alpha_{SATE} = \beta + \delta E(X | D = 1),$$

que pueden ser estimados fácilmente utilizando una regresión lineal.

- Alternativamente, se pueden tratar $\mu_1(X) = E(Y | D = 1, X)$ y $\mu_0(X) = E(Y | D = 0, X)$ como funciones no paramétricas de X , estimándolas mediante métodos kernel o de otro tipo.
- Esto último está más próximo al espíritu de la literatura de *matching*, que ha enfatizado las comparaciones directas y la ausencia de dependencia en supuestos de modelización.

4.3. Variables X discretas

- Supongamos que X toma J valores diferentes $\{\xi_1, \dots, \xi_J\}$ y tenemos una muestra de N observaciones X_1, \dots, X_N , de tal manera que cada una de las X_i ($i = 1, \dots, N$) es igual a uno de los J valores ξ_j ($j = 1, \dots, J$). Sea

N^j = número de observaciones en la celda j .

N_1^j = número de observaciones en la celda j con $D = 1$.

\bar{Y}_1^j = media de Y en la celda j para aquellos con $D = 1$.

\bar{Y}_0^j = media de Y en la celda j para aquellos con $D = 0$.

- Por tanto, $(\bar{Y}_1^j - \bar{Y}_0^j)$ es el equivalente muestral de $E(Y | D = 1, X = \xi_j) - E(Y | D = 0, X = \xi_j)$, a partir de lo cual podemos obtener las estimaciones

$$\hat{\alpha}_{ATE} = \sum_{j=1}^J (\bar{Y}_1^j - \bar{Y}_0^j) \frac{N^j}{N}$$

$$\hat{\alpha}_{SATE} = \sum_{j=1}^J (\bar{Y}_1^j - \bar{Y}_0^j) \frac{N_1^j}{N_1}$$

- La fórmula anterior también se puede escribir de la forma

$$\hat{\alpha}_{SATE} = \frac{1}{N_1} \sum_{D_i=1} (Y_i - \bar{Y}_0^{j(i)})$$

donde $j(i)$ indica la celda a la que pertenece X_i . Por tanto, $\hat{\alpha}_{SATE}$ empareja cada observación de una persona tratada con la media de los no tratados en la misma celda.

4.4. Variables X continuas

- Si el número de variables en X es pequeño, la idea anterior se puede extender al caso continuo:
bien discretizando el rango de valores de las X ,
bien emparejando observaciones de tratados y controles con valores parecidos de las X .

4.5. Métodos basados en el “propensity score”

- Rosenbaum y Rubin (Biometrika, 1983) denominaron *propensity score* a

$$\pi(X) = \Pr(D = 1 | X)$$

y demostraron que si $(Y_1, Y_0) \perp D | X$ entonces

$$(Y_1, Y_0) \perp D | \pi(X)$$

siempre que $0 < \pi(X) < 1$ para todo X .

- El resultado se prueba por simple aplicación de la ley de las esperanzas iteradas:

$$\begin{aligned} \Pr(D = 1 | Y_1, Y_0, \pi(X)) &\equiv E(D | Y_1, Y_0, \pi(X)) \\ &\equiv E[E(D | Y_1, Y_0, X) | Y_1, Y_0, \pi(X)] \\ &= E[\pi(X) | Y_1, Y_0, \pi(X)] \\ &= \pi(X) = \Pr(D = 1 | \pi(X)). \end{aligned}$$

- Respecto al último paso, nótese que

$$\begin{aligned} E[D | \pi(X)] &= E[E(D | X) | \pi(X)] \\ &= E[\pi(X) | \pi(X)] = \pi(X). \end{aligned}$$

- El interés de este resultado está en darse cuenta de que podemos emparejar personas con valores de X muy diferentes con tal de que tengan valores similares de $\pi(X)$.

- La implicación es que

$$\begin{aligned} &E[Y_1 - Y_0 | \pi(X) = \bar{\pi}] \\ &= E[Y | D = 1, \pi(X) = \bar{\pi}] - E[Y | D = 0, \pi(X) = \bar{\pi}] \end{aligned}$$

por lo que podemos emparejar utilizando el *propensity score* en vez del vector de X directamente.

- Alternativamente, podemos estimar de forma no paramétrica $E[Y | D = 1, \pi(X)]$ y $E[Y | D = 0, \pi(X)]$ tal como hacen Heckman, Ichimura y Todd (*Review of Economic Studies*, 1997).
- Los resultados anteriores sugieren procedimientos en dos etapas en los que se empieza por estimar el *propensity score*. No obstante, estos métodos implican decisiones sobre cómo modelizar $\pi(X)$ y cómo tenerlo en cuenta en la segunda etapa.

4.6. La condición de soporte común

- El conjunto de valores que toman las X para los tratados puede ser bastante distinto del de los no tratados. Heckman et al. (1997) encuentran que la violación de la condición de soporte común para las variables de emparejamiento (“comparing the incomparable”) es una fuente importante de sesgo en la evaluación.
- Se trata pues de restringir la aplicación de los métodos no experimentales a regiones de soporte común. Si S es el soporte común obtenemos

$$M(S) = \frac{\int_S E(Y_1 - Y_0 | X) dF(X | D = 1)}{\int_S dF(X | D = 1)}$$

o bien

$$M(S) = \frac{\int_S E(Y_1 - Y_0 | \pi(X)) dF(\pi(X) | D = 1)}{\int_S dF(\pi(X) | D = 1)}.$$

4.7. Incentivos monetarios y escolarización

- En septiembre de 1999 se puso en marcha en el Reino Unido el esquema piloto de un sistema de *Education Maintenance Allowance* (EMA) o prestaciones de sostenimiento educativo.
- Por medio del EMA se paga a los jóvenes entre 16 y 18 años que continúan estudios (tras completar los 11 grados de la educación obligatoria) una cantidad semanal entre 30 y 40 libras, más complementos por buenos resultados de hasta 140 libras a final de curso.
- La elegibilidad y en su caso la cuantía de los pagos depende de las circunstancias del hogar. Hogares con rentas anuales inferiores a £13000 (unos 3,3 millones de pesetas) son elegibles para los pagos máximos. No son elegibles aquellos con rentas superiores a £30000.
- Dearden, Emmerson, Frayne y Meghir (2002) participaron en el diseño del piloto y llevaron a cabo la evaluación de la política.
- No se llevó a cabo un diseño experimental a nivel de personas, sino que se definieron zonas de tratamiento (“zonas piloto”) y zonas de control, tanto rurales como urbanas.

- La pregunta básica que se plantea la evaluación es si se consigue más educación como resultado de la política.
- La preocupación es que algunas familias no tomen decisiones óptimas como resultado de restricciones de liquidez o de información.
- Utilizan técnicas basadas en el *propensity score*. Para obtener $\pi(X)$, Dearden et al. estiman un modelo probit con gran número de características familiares, locales y de escuela (y sus interacciones).
- Para cada observación en un área de control construyen un contrafactual promedio utilizando una regresión *kernel* y calculan los errores estándar utilizando *bootstrap*.
- Encuentran que EMA aumentó de forma significativa la participación en el año 12 (el primero no obligatorio), en un 5.9% para los elegibles y en un 3.7% para toda la población.
- Sólo se encontraron efectos significativos para aquellos que recibieron pagos completos.

5. Variables instrumentales

- Supongamos que tenemos datos no experimentales y que las X disponibles nos impiden hacer un supuesto de independencia condicional como en los métodos de *matching*:

$$(Y_1, Y_0) \perp D \mid X.$$

- Supongamos no obstante que tenemos una variable Z que proporciona una “fuente de variación exógena en D ” en el sentido de que se cumple el *supuesto de independencia*:

$$(Y_1, Y_0) \perp Z \mid X$$

y a la vez el *supuesto de relevancia*:

$$Z \text{ dep. } D \mid X.$$

- En un ejemplo clásico, Z es el indicador de asignación a grupos de tratamiento y control en un diseño experimental. Por lo tanto $(Y_1, Y_0) \perp Z$. Sin embargo, el indicador real de tratamiento D difiere de Z porque algunas personas del grupo de tratamiento deciden no seguirlo (*non-compliers*). Claramente, aunque Z y D no coinciden, en general no serán independientes.
- La cuestión es si nos sirve esta situación para medir algún tipo de efecto causal. Para responder a esta pregunta considero el caso en que Z es binaria. Además, hacemos abstracción del hecho de que la argumentación puede ser condicional a X .

5.1. Efectos homogéneos

- Si el efecto causal es el mismo para todos los individuos

$$Y_{1i} - Y_{0i} = \alpha$$

entonces tener una variable instrumental nos permite identificar α . Esta es la situación tradicional en modelos econométricos con variables explicativas endógenas. En general

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) D_i$$

y en el caso homogéneo

$$Y_i = Y_{0i} + \alpha D_i.$$

Además, teniendo en cuenta que $Y_{0i} \perp Z_i$

$$E(Y_i | Z_i = 1) = E(Y_{0i}) + \alpha E(D_i | Z_i = 1)$$

$$E(Y_i | Z_i = 0) = E(Y_{0i}) + \alpha E(D_i | Z_i = 0).$$

Restando ambas ecuaciones obtenemos

$$\alpha = \frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(D_i | Z_i = 1) - E(D_i | Z_i = 0)}$$

lo que sirve para determinar α siempre que

$$E(D_i | Z_i = 1) \neq E(D_i | Z_i = 0).$$

- Intuitivamente el efecto de D sobre Y se puede medir a través del efecto de Z porque hemos supuesto que Z sólo afecta a Y a través de D .

5.2. Efectos heterogéneos

- En el caso heterogéneo la existencia de variables instrumentales no es suficiente para identificar efectos causales.

- Un supuesto adicional que permite identificar α_{SATE} es una regla de elegibilidad de la forma

$$\Pr(D = 1 | Z = 0) = 0$$

i.e. a las personas con $Z = 0$ se les deniega el tratamiento.

- Otro supuesto adicional es la siguiente condición de “monotonidad”: Cualquier persona que estuviese dispuesta a seguir el tratamiento habiendo sido asignada al grupo de control también estaría dispuesta a seguirlo en caso de ser asignada al grupo de tratamiento.
- La plausibilidad de este supuesto depende del contexto de aplicación.
- En tal caso, el coeficiente de variables instrumentales coincide con el efecto medio del tratamiento para aquellas personas cuyo valor de D cambia cuando cambia su valor de Z (*local average treatment effect* o LATE).

5.3. Barrios étnicos y éxito de los inmigrantes

- Interés en las consecuencias para el éxito laboral de vivir en barrios de alta concentración étnica o nacional. En Suecia el 11% de la población ha nacido en el extranjero. De estos, más de un 40% viven en un barrio étnico.
- En principio, el efecto causal de vivir en un barrio étnico para los inmigrantes es ambiguo. Por un lado, la segregación residencial puede reducir la tasa de adquisición de conocimientos propios del país de adopción (eg. la lengua) y de ahí dificultar el acceso a mejores empleos. Por otro, el enclave puede representar una red que aumenta las oportunidades de sus miembros diseminando información a nuevos inmigrantes.
- Si se comparan directamente, los ingresos de los inmigrantes que viven en barrios étnicos son un 5% inferiores a los que no, teniendo en cuenta diferencias en edad, estudios, sexo, situación familiar, lugar de procedencia y año de inmigración.
- El problema es que esta asociación estadística puede no tener valor causal si la decisión de vivir en un barrio étnico está asociada con las oportunidades esperadas.

- Sea el ingreso de un inmigrante si vive en un barrio étnico

$$Y_1 = \mu_1(Z) + v_1$$

y sea el ingreso de la misma persona si decide no vivir en un barrio étnico

$$Y_0 = \mu_0(Z) + v_0.$$

- La ganancia (o pérdida) por vivir en un barrio étnico para esta persona es

$$Y_1 - Y_0$$

y la ganancia media para todas las personas con características Z es

$$b(Z) = \mu_1(Z) - \mu_0(Z)$$

- El ingreso observado es

$$Y = \begin{cases} Y_1 & \text{si } X = 1 \\ Y_0 & \text{si } X = 0 \end{cases}$$

Por tanto,

$$\begin{aligned} \alpha(Z) &= E(Y | Z, X = 1) - E(Y | Z, X = 0) \\ &= E(Y_1 | Z, X = 1) - E(Y_0 | Z, X = 0) \\ &\neq E(Y_1 | Z) - E(Y_0 | Z) \end{aligned}$$

a menos que X sea independiente de Y_1, Y_0 dado Z .

- El análisis de regresión nos dice que $E[\alpha(Z)]$ está entorno al -5% . Solo si Z fuese suficientemente preciso podríamos dar credibilidad causal a esta cifra.

- Entre 1985 y 1991 el gobierno sueco asignó un lugar de residencia inicial a inmigrantes refugiados. La motivación de la política era que la dispersión de inmigrantes facilitaría su integración.
- Sea W un indicador de la asignación inicial (8 años antes de la medición de X). Edin, Fredriksson y Aslund (2002) suponen que W es independiente de v_0 y v_1 , pero está asociado con X .
- Las estimaciones de variables instrumentales realizadas por Edin et al. (2002) dan lugar a una ganancia para inmigrantes de baja cualificación del 13% asociada con el aumento de una desviación estándar en la concentración étnica.
- Para los inmigrantes del alta cualificación no se observa ni ganancia ni pérdida.

5.4. Comentarios finales

- Cuando la acción de una política crea una situación de este tipo se habla de *experimento natural*, en el sentido de que se crea el tipo de variación exógena propio de los experimentos controlados.
- El método de variables instrumentales está en la base de la teoría de ecuaciones simultáneas desarrollada por los economistas de los años 40 y 50.
- En el marco clásico de ecuaciones simultáneas el objetivo es determinar una estructura. Por el contrario, en el ejemplo anterior se utilizan variables instrumentales para determinar un *efecto causal de forma reducida* (resultado de interacciones entre distintos tipos de efectos).
- La situación es similar a la que es común en las ciencias naturales: eg. mediante un experimento controlado podemos establecer una relación causal entre el tabaco y las enfermedades coronarias. Se trata sin embargo de un efecto causal de forma reducida. El mecanismo a través del cual el tabaco actúa sobre la propensión a la enfermedad no es explicado.