

Econometría de evaluación de impacto

LUIS GARCÍA NÚÑEZ

Pontificia Universidad Católica del Perú

RESUMEN

En años recientes los métodos de evaluación de impacto se han difundido ampliamente en la investigación microeconómica aplicada. Sin embargo, la variedad de métodos responde a problemas particulares y específicos los cuales están determinados normalmente por los datos disponibles y el impacto que se busca medir. El presente documento resume las principales corrientes disponibles en la literatura actual, poniendo énfasis en los supuestos bajo los cuales el efecto tratamiento promedio y el efecto tratamiento promedio sobre los tratados se encuentran identificados. Adicionalmente se presentan algunos ejemplos de aplicaciones prácticas de estos métodos. Se busca hacer una presentación didáctica que pueda ser útil a estudiantes avanzados y a investigadores aplicados que busquen conocer los principios básicos de estas técnicas.

Palabras clave: inferencia causal, evaluación de programas, regresión discontinua, variables instrumentales, *matching*

Clasificación JEL: C13, C14, C31

ABSTRACT

In recent years the program evaluation methods have become very popular in applied microeconomics. However, the variety of these methods responds to specific problems, which are normally determined by the data available and the impact the researcher tries to measure. This paper summarizes the main methods in the current literature, emphasizing the assumptions under which the average treatment effect and the average treatment effect on the treated are identified. Additionally, after each section I briefly present some applications of these methods. This document is a didactic presentation for advanced students in economics and applied researchers who wish to learn the basics of these techniques.

Keywords: causal inference, program evaluation, regression discontinuity, instrumental variables, *matching*

JEL Classification: C13, C14, C31

1. INTRODUCCIÓN Y EL ANÁLISIS CAUSAL EN ECONOMÍA

En décadas recientes, los estudios de evaluación de impacto de políticas públicas se han realizado con métodos estadísticos y econométricos cada vez más sofisticados, con el fin de obtener una evaluación científicamente rigurosa basada en *el análisis de causalidad*. La popularidad de estos métodos ha llevado a que se busque aplicarlos en numerosos contextos. Sin embargo, por lo general, se basan en ciertos supuestos poblacionales que condicionan su radio de acción y que definen hasta dónde se pueden obtener conclusiones valederas (léase *causales*) de las evaluaciones.

El interés por estudiar las relaciones causales en el mundo real ha sido crucial en la ciencia a lo largo de su historia. El concepto de causalidad ha ido variando a lo largo de los siglos. Sin embargo, el fondo de la discusión sigue siendo el mismo: el interés por hacer un estudio acerca de la relación causal entre las variables. Este estudio empieza con la pregunta inicial de cualquier estudio de impacto: ¿cuál es el efecto causal de una variable *X* sobre otra variable *Y*? Responderla puede ser un asunto no tan trivial desde el punto de vista analítico ni desde los datos. Pues para tener una idea de este efecto, deberíamos tener alguna idea sobre la existencia de una relación causal entre estas variables.

Desde el punto de vista de la estadística, según Holland (1986), durante mucho tiempo la estadística no se preocupó de los temas causales, limitándose a señalar que «la correlación no implica causalidad». Esto se debió a que tradicionalmente la estadística inferencial ha estudiado la manera cómo los datos «aparecen» en el mundo real. Tal interés conlleva al estudio de la distribución de probabilidad conjunta de estas variables, la cual entrega las probabilidades de ocurrencia de ellas, y la obtención de «parámetros asociativos», tales como las probabilidades y esperanzas condicionales. Estos parámetros no son determinantes para establecer relaciones causales entre las variables. La presencia de variables asociadas sin mayor sentido, como en el caso de las conocidas correlaciones espurias o la presencia de los llamados *confounders*, presenta una limitación importante para el análisis de inferencia causal basándose en parámetros asociativos cuando tales *confounders* no son tomados en cuenta.

Desde la visión de la economía y la econometría, desde las primeras décadas del siglo XX, el análisis de causalidad se configuró como el estudio de las variables del mundo real, estableciendo algún tipo de ordenamiento secuencial o lógico entre ellas (Goldberger 1972). De esta manera, y bajo supuestos teóricos o de juicio no «testeables»¹, se pues establecer una estructura de ramificaciones causales que une a aquellas variables y que generan los datos observados.

Cuando aplicamos este análisis a la economía, encontramos que el proceso generador de datos está gobernado por relaciones económicas subyacentes a él (véase por ejemplo, Haavelmo 1943, 1944). Estas relaciones suelen ser simplificadas y sistematizadas a través

¹ Un ejemplo de un supuesto causal que no necesita ser verificado es que ninguna variable puede ocasionar un cambio en la edad de las personas.

de los llamados «modelos económicos», los cuales definen claramente a sus variables exógenas y a sus endógenas, y cuyas relaciones causales son el resultado del comportamiento de los agentes económicos (por ejemplo, la maximización de ganancias de la empresa)². En estos modelos, las variables exógenas tienen efecto sobre las endógenas, y no al revés, y por ello podemos afirmar que las relaciones de causalidad entre variables económicas tienen en sí mismas un sustento en la teoría económica. Asimismo, el impacto de una variable sobre otra puede ser estudiado mediante una alteración de una de las variables exógenas y observando su efecto sobre una o más variables endógenas, manteniendo todo lo demás constante (la noción marshalliana de un cambio *ceteris paribus*).

En el terreno de los datos económicos poblacionales, estos no ocurren por el mero azar sino que aparecen por relaciones estructurales entre las variables mencionadas arriba. Estas estructuras no se limitan solamente a las variables observables consideradas en el modelo económico sino que también incluyen a aquellas que no son observables pero que suelen tener un rol importante en la estructura (Pearl 2000, 2009; Heckman 2000, 2005). También se consideran mecanismos de selección de información intrínsecos al problema económico bajo estudio que tampoco ocurren al azar³. Estas relaciones pueden ser escritas en forma de ecuaciones, con lo cual se definen los modelos de *ecuaciones estructurales*. En tales ecuaciones se representan relaciones causales y no meras asociaciones empíricas.

El análisis causal basado en ecuaciones estructurales es el más completo pues ofrece una visión panorámica del conjunto, lo cual permite entender especialmente las dificultades que pueden surgir en el proceso de análisis del efecto de una variable sobre otra. No obstante, este enfoque ha recibido críticas debido a que sus conclusiones pueden depender muy sensiblemente de los supuestos sobre los términos de perturbación (Angrist, Imbens y Rubin 1996). Heckman (2000, 2005), aunque defiende al enfoque estructural, señala que los principales cuestionamientos que dicho enfoque ha recibido se basan en deficiencias de estos modelos para reproducir los datos observados.

En términos empíricos, algunos estudios (por ejemplo, Lalonde 1986) han comprobado que empíricamente los modelos estructurales basados en regresiones entregan pobres resultados en comparación con métodos más «experimentales». Asimismo, pueden sufrir problemas de error de especificación, los cuales suelen ser magnificados en modelos multiecuacionales. Por último, pueden ser complicados, y sobre todo poco práctico si el objetivo es analizar el impacto entre dos variables y si no estamos muy interesados en estudiar a profundidad al resto de variables que las circundan.

Es quizás por esta razón que recientemente se ha incrementado la popularidad de los estudios de impacto basados en métodos experimentales y cuasiexperimentales en la

² Cabe mencionar que más de un modelo económico podría ser consistente con los mismos datos.

³ Por ejemplo, la participación laboral de las mujeres es un típico ejemplo de un mecanismo de autoselección, el cual trunca la información sobre los salarios observados de las mujeres trabajadoras.

investigación aplicada en economía, en similitud con estudios equivalentes en estadística y epidemiología, por ejemplo. Tales trabajos buscan en principio encontrar el efecto de una política, aunque descuidan notablemente el fundamento económico y algunos aspectos econométricos (como la identificación, los mecanismos de selección y el comportamiento de las variables no observables).

Puesto que la mira de estas metodologías se encuentra en los experimentos, necesitaremos un armazón analítico simple a seguir. Este es el modelo de resultados potenciales, el cual en economía se remonta a Roy (1951) y Quandt (1958), y que Maddala (1983) presenta con el nombre de *switching regression model*, pero que en estadística se le conoce como el modelo de Neyman-Rubin, por los trabajos de Neyman (1990) y Rubin (1974). En un principio presentaremos este modelo, para luego poco a poco ir estudiando los principales problemas de identificación que pueda tener en diferentes contextos, proponiendo estrategias identificadoras del efecto causal buscado.

Este modelo tiene sus fundamentos en los modelos de ecuaciones estructurales, aunque su presentación es más simple al basarse en los estudios experimentales, teniendo al experimento aleatorio controlado como su paradigma. Se trata de aislar el efecto de una variable (que podría ser de política) sobre una variable endógena, manteniendo cualquier otro factor que la afecte de manera «controlada» o constante (*ceteris paribus*). Luego se calcula la diferencia en la variable endógena en los dos estados, con y sin la aplicación de la política, entendiéndose tal diferencia como el efecto causal de la política. Sin embargo, y tal como quedará claro en la siguiente sección, tal efecto no puede ser observado debido a que solo uno de los resultados potenciales es observable. Existe entonces un serio problema de identificación, llamado el problema fundamental de la inferencia causal.

Normalmente al escenario no observable se le llama «contrafactual» o «hipotético». Se podría pensar que el escenario contrafactual puede ser observado si, por ejemplo, sometemos a la política a un individuo que originalmente no participó en este programa. No obstante, eso no es cierto pues no se estaría cumpliendo la condición de *ceteris paribus* porque al menos alguna cosa debió cambiar en el tiempo⁴.

⁴ En algunos experimentos podría creerse que se puede conocer ambos estados de la naturaleza, por ejemplo, encender y apagar la luz para ver el efecto de la corriente eléctrica en un bombillo de luz. En este ejemplo es casi seguro que cualquier otro factor que afecte la luminosidad del bombillo está bajo control del investigador y, por lo tanto, el escenario «antes del tratamiento» y «después del tratamiento» pueden ser considerados como los dos resultados potenciales. En general no ocurre lo mismo en otros estudios, en donde los dos escenarios mencionados no necesariamente mantienen constantes a los demás factores que podrían afectar a la variable *y*. Por ejemplo, el efecto de la lactancia materna sobre la incidencia de enfermedades en los infantes no puede ser estudiado observando simplemente el «antes» y el «después» de la exposición al tratamiento pues existen factores que cambian en forma natural (como la edad y el peso del niño) y además otros factores podrían cambiar circunstancialmente (como las condiciones de vida de la familia), a pesar que algunos factores sí se mantengan constantes (como el sexo del niño y su resistencia natural a las enfermedades).

Afortunadamente, los orígenes del problema han sido estudiados y entendidos, y por lo tanto somos capaces de proveer soluciones a él. Tales estrategias se basan en la aplicación de supuestos y además con un importante apoyo de la estadística, se logra identificar el efecto causal. Este documento se basa justamente en estas estrategias de identificación.

Sin embargo, este no es el único problema de identificación que enfrenta el análisis causal econométrico de políticas. En este documento se revisan algunas de las técnicas más populares, poniéndose énfasis en las condiciones bajo las cuales se puede identificar a los efectos causales.

En la sección 2 se presentan algunas cuestiones básicas sobre el lenguaje de evaluación de impacto. En la sección 3 se presentan a los experimentos aleatorios controlados (paradigma usado en muchas disciplinas para estudiar efectos causales) y los estudios no experimentales (que son los más usados en economía). En la sección 4 se explica el diseño de regresión discontinua aguda, que luego se retoma en la sección 7 con el diseño de regresión discontinua difusa. En la sección 5 se explica el método de pareo o *matching*. En la sección 6 se presenta el método de variables instrumentales aplicado a la evaluación de impacto de políticas. En la sección 8 se presenta brevemente el método de diferencias en diferencias. Finalmente en la sección 9 se presenta algunas reflexiones y consideraciones finales.

2. ALGUNAS CUESTIONES BÁSICAS

2.1. EL MODELO DE RESULTADOS POTENCIALES Y EL EFECTO TRATAMIENTO PROMEDIO (ATE)⁵

Con el fin de estudiar la identificación del efecto causal, formalicemos lo expuesto anteriormente del modelo de resultados potenciales concentrándonos en un caso especial. Supongamos que deseamos conocer el efecto de un tratamiento d (por ejemplo, una política) sobre alguna variable de interés y_i (un resultado), para $i = 1, \dots, N$, donde i indica una unidad i .

Por ejemplo:

Tratamiento (d)	Resultado (y)
Ejercicio diario	Presión sanguínea
Capacitación laboral	Salarios
Un nuevo reglamento de tránsito	Tasa de accidentes de tránsito
Un medicamento	Colesterol

⁵ La notación y definiciones que seguimos en esta sección está influenciada en la exposición de Lee (2005).

Aunque el tratamiento podría ser en diferentes intensidades, y al mismo tiempo los resultados podrían ser múltiples, vamos a simplificar el análisis considerando que el tratamiento d es binario, tomando el valor 1 si la unidad recibe el tratamiento y 0 si no la recibe.

$$d_i = \begin{cases} 1 & \text{Si } i \text{ recibe el tratamiento} \\ 0 & \text{Si } i \text{ no lo recibe} \end{cases}$$

Tenemos una población U de unidades, algunas de las cuales recibirá un tratamiento. Cada unidad $i \in U$ puede ser descrita por el siguiente conjunto $(y_{0i}, y_{1i}, d_i, x_i, \varepsilon_i)$ donde:

y_{0i} = resultado potencial si la unidad i no recibió el tratamiento

y_{1i} = resultado potencial si la unidad i recibió el tratamiento

x_i = vector de características observables de la unidad i

ε_i = vector de características no observables de la unidad i

Cabe mencionar que la condición de observable o no observable de las características se define desde el punto de vista del investigador o evaluador de la política.

Definamos el resultado observado y_i como $y_i = d_i y_{1i} + (1 - d_i) y_{0i}$, el cual es igual a uno de los resultados potenciales. Asimismo podemos clasificar a todas las unidades de la población según la recepción o no del tratamiento. Como nos preocupa analizar el impacto de políticas (micro) económicas, llamaremos a los receptores de la política como el grupo beneficiario, definido como $B = \{i \in U / d_i = 1\}$. Al grupo de unidades que no recibe el tratamiento lo llamaremos grupo no beneficiario⁶ $N = \{i \in U / d_i = 0\}$.

Lo único que podemos observar para una unidad $i \in U$ es el paquete $(y_{1i}, x_i, d = 1)$ y para una unidad $k \in U$ en el grupo no beneficiario solo observamos $(y_{0k}, x_k, d = 0)$.

El efecto tratamiento individual para una unidad i , $\delta_i = y_{1i} - y_{0i}$, no está identificado pues uno de sus elementos no es observable. Sin embargo, podría ser más conveniente analizar el efecto tratamiento promedio para la población (*ATE* por sus siglas en inglés). Omitiendo el subíndice i , el *ATE* es el parámetro poblacional

$$\delta = ATE = E(y_1 - y_0) = E(y_1) - E(y_0)$$

Debido a que los valores potenciales y_0 y y_1 no son plenamente observables para todo $i \in U$, se debe tener cuidado al estimar este valor esperado usando análogos de muestras, como el promedio simple por ejemplo.

⁶ En algunos estudios se le llama también grupo de control a aquel que no ha recibido el tratamiento. Sin embargo dado que el énfasis en este estudio recae en los llamados estudios observacionales (véase sección 3.2) en donde los datos disponibles no provienen de experimentos controlados, conviene llamar a este grupo simplemente como no beneficiario, reservando el nombre de «grupo de control» para aquel grupo que no recibe tratamiento en estudios experimentales o también a un subgrupo de los no beneficiarios que cumplen ciertas características (que se discutirán más adelante, véase la sección 3.1) en estudios observacionales.

$$\hat{\delta} = \frac{1}{n_B} \sum_{i \in B} y_i - \frac{1}{n_N} \sum_{i \in N} y_i = \bar{y} |_{i \in B} - \bar{y} |_{i \in N}$$

Donde n_B es el número de beneficiarios y n_N es el número de no beneficiario.

El peligro de cometer un error con una estimación de esta manera se basa en el conocido «problema de la selección»: la no observación de los valores de y_{0i} y y_{1i} para algunos individuos podría responder a una conducta sistemática de los individuos o de los otorgantes del beneficio. Por ejemplo, si se busca analizar el efecto de la participación en programas de ejercicios físicos en el estado de salud medido como los niveles de presión sanguínea, es claramente factible que aquellos que finalmente acepten participar en el tratamiento sean individuos que tengan fuertes preferencias por la actividad física, o que presenten determinadas características como su edad y peso. Por el contrario, aquellos que opten por no participar en el programa podrían haber tomado esta decisión basándose en las mismas características o preferencias. En este ejemplo estaría ocurriendo un problema de «autoselección» en el tratamiento, donde la participación en el programa dependerá de las características observables de las personas (su edad y peso) o de características no observables (sus preferencias, hábitos de vida, factores genéticos, etc.)

La selección podría haber venido por parte de los diseñadores de la política. Por ejemplo, si fijan una población objetivo para el tratamiento o si priorizan a algunos grupos que ya de por sí presenten problemas de presión arterial, nuevamente existirían diferentes características (observables o no) en los grupos B y N .

Si esto es lo que está ocurriendo con el programa d , entonces el estimador propuesto $\hat{\delta}$ es el análogo muestral de $E(y | d = 1) - E(y | d = 0)$, el cual es en general diferente de $E(y_1 - y_0)$, cuando las características (x_i, ε_i) difieren entre los beneficiarios y no beneficiarios. Por esta razón se debe analizar con cuidado (a) en qué casos $\hat{\delta}$ es un buen estimador de ATE ; (b) qué otro estimador distinto de $\hat{\delta}$ podría estimar correctamente a δ .

2.2. SUPUESTOS IDENTIFICADORES DEL ATE

Supongamos que el tratamiento o política ha sido aplicado a los individuos de una manera muy particular. Digamos que se ha realizado un sorteo en donde cada individuo tiene la misma probabilidad de recibir el beneficio. En tal caso, el tratamiento d será independiente de los resultados potenciales y_j , para $j = 0, 1$. Formalmente diremos:

(I): Los resultados potenciales son estadísticamente independientes de d . En símbolos $(y_0, y_1) \perp d$.

Dada esta condición de independencia, entonces ocurrirá que

$$\begin{aligned} ATE &= \delta = E(y_1 - y_0) = E(y_1) - E(y_0) = E(y_1 | d = 1) - E(y_0 | d = 0) \\ &= E(y | d = 1) - E(y | d = 0) \end{aligned}$$

La última igualdad ocurre porque y_1 solo es observable cuando $d = 1$, con ello coinciden y_1 con y , y lo mismo ocurre para y_0 . Por lo tanto si tal supuesto se cumple, entonces *ATE* puede ser estimado consistentemente simplemente con la diferencia de los promedios simples de las observaciones de los grupos B y N , o sea el estimador $\hat{\delta}$. Nótese que este estimador es igual al estimador $\hat{\beta}_d$, que se obtendría de la estimación por mínimos cuadrados ordinarios del modelo de regresión lineal $y_i = \beta_1 + \beta_d d_i + u_i$.

No es necesario un supuesto tan fuerte, como el de independencia, para que se cumpla este resultado. Una condición más débil que es implicada por el supuesto de independencia es la siguiente:

(II): y_0 y y_1 son «independientes en medias» de d si $E(y_j | d) = E(y_j)$, para $j = 0, 1$. De forma equivalente, $E(y_j | d = 1) = E(y_j | d = 0)$.

Bajo esta condición se cumple también que el *ATE* coincide con la diferencia $E(y | d = 1) = E(y | d = 0)$.⁷

2.3. EL EFECTO TRATAMIENTO SOBRE LOS TRATADOS (*ATET*)

Es frecuente que los programas no tengan aplicabilidad universal sino solamente en parte de la población. Por ejemplo, un programa de desempleo solo interesa en la población de desempleados, no toma en cuenta a los empleados. En tal caso, el impacto del programa se mide únicamente en el grupo tratado, pues nos interesa comparar la situación real del grupo beneficiario con la situación contrafactual de ellos mismos en el caso hipotético de que no hubieran recibido el beneficio del programa, sin importarnos mucho el efecto sobre los no tratados. A este impacto se le llama el Efecto Tratamiento Promedio en los Tratados⁸ o *ATET*:

$$\delta_T = ATET = E(y_1 - y_0 | d = 1) = E(y_1 | d = 1) - E(y_0 | d = 1)$$

Con la información disponible, el primer término de *ATET* es plenamente observable pues es solamente la esperanza condicional del resultado dado que los individuos participaron en el programa, es decir $E(y | d = 1)$. En cambio, el segundo término no es observable pues no disponemos de información del resultado potencial y_0 cuando $d = 1$, existiendo por lo tanto un problema de identificación de este parámetro.

Este término será identificable si se supone que $y_0 \perp d$ (o con el supuesto más débil de y_0 independiente en media de d , $E(y_0 | d = 1) = E(y_0 | d = 0)$). En tal caso se puede estimar el segundo componente de *ATET* con un análogo muestral de $E(y | d = 0)$. En términos intuitivos, este supuesto quiere decir que el tratamiento ha sido asignado entre

⁷ Es frecuente hacer explícito un supuesto adicional: el supuesto de no interferencia entre las unidades bajo estudio. Según este supuesto, el tratamiento aplicado a una entidad solo le afecta a ella misma y no a nadie más. En inglés se le conoce como SUTVA (*Stable Unit Treatment Value Assumption*), el cual se asumirá implícitamente a lo largo de este documento.

⁸ En inglés es el *Average Treatment Effect on the Treated*.

los individuos de los grupos de beneficiarios y no beneficiarios, independientemente del resultado potencial que ellos hubieran obtenido sin tratamiento y_0 . Sin embargo, es posible que y_1 no sea independiente de d , lo cual no afectaría la identificación de ATE . Por ejemplo, y_1 no sería independiente de d si los individuos participantes se autoseleccionan para participar en el programa porque tendrían una ganancia esperada de y_1 más alta que aquellos que no participan.

En general se cumplirá que ATE es distinto de ATE . Sin embargo, podrían ser exactamente iguales si se cumple ya sea los supuestos (I) o (II). Para mostrar esto,

$$\begin{aligned}ATE &= E(y_1 - y_0 | d = 1) = E(y_1 | d = 1) - E(y_0 | d = 1) = E(y_1 | d = 1) - E(y_0 | d = 0) \\ &= E(y | d = 1) - E(y | d = 0) = ATE\end{aligned}$$

Entonces, al ser ATE y ATE iguales bajo este supuesto, ambos pueden ser estimados mediante el estimador $\hat{\delta}$, que es la diferencia de los promedios simples de los grupos de beneficiarios y no beneficiarios.

Por otro lado, trivialmente ATE y ATE también podrían ser iguales si el programa se aplicara a toda la población.

$$\begin{aligned}E(y_1 - y_0) &= E(y_1 - y_0 | d = 1) \cdot \Pr(d = 1) + E(y_1 - y_0 | d = 0) \cdot \Pr(d = 0) \\ ATE &= ATE \cdot \Pr(d = 1) + ATEU \cdot \Pr(d = 0)\end{aligned}$$

Donde $ATEU$ es el efecto tratamiento sobre los no tratados (un parámetro de escaso interés práctico). Luego si $\Pr(d = 1) = 1$ tendríamos $ATE = ATE$, lo cual equivaldría a aplicar el programa a toda la población.

2.4. CONDICIONAMIENTO A CARACTERÍSTICAS OBSERVABLES

Los resultados mencionados se pueden generalizar si se condicionan a las características observables x , lo que podría entenderse como limitar el análisis a una subpoblación con características x . Por ejemplo, se podría calcular el efecto tratamiento promedio según el sexo de la persona o su nivel educativo, su estado civil, etc.

Las definiciones de ATE y ATE con condicionamiento a x son: $ATE|x = E(y_1 - y_0 | x)$ y $ATE|x = E(y_1 - y_0 | d = 1, x)$. En tal caso, los supuestos identificadores de estos parámetros se generalizan como:

(I) y_j es estadísticamente independiente de d , dado x : $y_j \perp d | x$.

(II) y_0 y y_1 son «independientes en media condicional» de d dado x : $E(y_j | d, x) = E(y_j | x)$, para $j = 0, 1$.

Cuando se condiciona por x , es frecuente hacer un supuesto adicional sobre la existencia de individuos beneficiarios y no beneficiarios para cada subpoblación x . A este supuesto se le conoce como supuesto de *matching* u *overlapping*.

$$(III) 0 < P(d = 1 | x) < 1$$

Luego, bajo los supuestos (II') y (III), el $ATE | x$ es igual a la diferencia de la media condicional de los grupos B y N .

$$\begin{aligned} E(y_1 - y_0 | x) &= E(y_1 | x) - E(y_0 | x) = E(y_1 | d = 1, x) - E(y_0 | d = 0, x) = \\ &= E(y | d = 1, x) - E(y | d = 0, x) \end{aligned}$$

Nótese que el último signo igual de la ecuación anterior no se cumpliría si no se efectuara el supuesto (III). Luego, el $ATE | x$ puede ser calculado como la diferencia simple de los promedios de y dado d para un subgrupo específico x .

Un resultado adicional que vale la pena mencionar en esta sección es que si asumimos que el tratamiento se asigna completamente al azar (mediante un sorteo simple), entonces el tratamiento d será también independiente de las características observables y no observables de los individuos (x_i, ε_i) , las cuales se encontrarán «balanceadas» entre los grupos B y N .

2.5. SESGO DEBIDO A LA VIOLACIÓN DE LOS SUPUESTOS

Cuando los supuestos mencionados antes no se cumplen, entonces el estimador propuesto $\hat{\delta}$ será sesgado al querer estimar a ATE o a ATE_T . El llamado problema de la selección presentado en la sección 2.1 provocará que existan sesgos cuyas fuentes estriban en el «desbalance» existente en las características observables y no observables entre los grupos B y N . Así, tanto el cálculo de estos parámetros incluirá a las diferencias de estas características.

Cuando los grupos B y N difieren en las características observables x , diremos que tenemos «selección en observables», mientras que si difieren en las variables no observables ε , tenemos «selección en no observables». En el primer caso, el sesgo sobre la diferencia de medias originado por la selección se llama en la literatura inglesa como *overt bias*, mientras que en el segundo, el sesgo se llama *covert bias* o *hidden bias*.

Formalmente las definiremos ambos tipos de selecciones así:

Selección en observables: $E(y_j | d) \neq E(y_j)$ pero $E(y_j | d, x) = E(y_j | x)$

Selección en no observables: $E(y_j | d, x) \neq E(y_j | x)$ pero $E(y_j | d, x, \varepsilon) = E(y_j | x, \varepsilon)$

Concentrándonos en la selección en observables, no se cumple el supuesto (II) pero si controlamos por las variables que ocasionan la selección, tendremos entonces que el ATE es identificable condicionado a un grupo particular x , pues se cumpliría el supuesto (II'), de la forma como se mencionó en la sección anterior.

Por el contrario, si tenemos selección en no observables, la condición en x no garantiza que la diferencia de las medias de grupo refleje el impacto del programa. Habría entonces que condicionar también en las características no observables para que $\hat{\delta}$ sea un estimador de ATE entendiendo esto como la definición de un subgrupo que comparta

las mismas características no observables. Esto es en la práctica difícil al ser justamente las características ε invisibles para el investigador.⁹

Podemos observar bajo qué condiciones el desbalance de las características x podría sesgar la estimación de ATE mediante la diferencia de medias simple $\hat{\delta}^{10}$. Supongamos que solo existe una característica x la cual es binaria, tomando el valor de 1 para algunos individuos y 0 para el resto. Supongamos también que se cumple el supuesto (II'). En este contexto, los efectos tratamiento promedio condicionados a x son

$$ATE|_{x=0} = E(y_1 | d = 1, x = 0) - E(y_0 | d = 0, x = 0) = E(y | d = 1, x = 0) - E(y | d = 0, x = 0)$$

$$ATE|_{x=1} = E(y_1 | d = 1, x = 1) - E(y_0 | d = 0, x = 1) = E(y | d = 1, x = 1) - E(y | d = 0, x = 1)$$

Por otro lado, podemos expresar a la diferencia de medias de y entre B y N como:

$$E(y | d = 1) - E(y | d = 0) = E(y | d = 1, x = 0) \cdot \Pr(x = 0 | d = 1) \\ + E(y | d = 1, x = 1) \cdot \Pr(x = 1 | d = 1) - E(y | d = 0, x = 0) \cdot \Pr(x = 0 | d = 0) \\ - E(y | d = 0, x = 1) \cdot \Pr(x = 1 | d = 0)$$

Reemplazando las expresiones de $ATE|_{x=0}$ y $ATE|_{x=1}$ en la expresión anterior tenemos:

$$E(y | d = 1) - E(y | d = 0) = ATE|_{x=0} \cdot \Pr(x = 0 | d = 1) + ATE|_{x=1} \cdot \Pr(x = 1 | d = 1) \\ + E(y | d = 0, x = 0) \cdot [\Pr(x = 0 | d = 1) - \Pr(x = 0 | d = 0)] \\ - E(y | d = 0, x = 1) \cdot [\Pr(x = 1 | d = 0) - \Pr(x = 1 | d = 0)] \quad (\clubsuit)$$

Si se cumpliera que $x \perp d$, entonces se cumpliría que $\Pr(x | d) = \Pr(x)$ para cualquier combinación de x y d , con lo cual:

$$E(y | d = 1) - E(y | d = 0) = ATE|_{x=0} \cdot \Pr(x = 0) + ATE|_{x=1} \cdot \Pr(x = 1) = ATE \quad (\spadesuit)$$

No obstante, el desbalance del x entre los beneficiarios y no beneficiarios provoca que el valor del lado derecho de la ecuación (\clubsuit) sea diferente del efecto tratamiento promedio mostrado en (\spadesuit).

Observando con cuidado el lado derecho de la ecuación (\clubsuit) notamos que los dos primeros términos están multiplicados por las probabilidades condicionales de x dado d , los cuales no necesariamente serán iguales a las probabilidades incondicionales de x debido al desbalance mencionado. Esto trae como consecuencia una primera desviación con respecto al efecto tratamiento promedio que queremos calcular. No obstante, este error desaparecería en el caso particular que $ATE|_{x=0} = ATE|_{x=1}$.

⁹ Aunque no podamos condicionar en no observables, mediante algunos procedimientos experimentales podemos confiar que, estadísticamente hablando, las características no observables puedan balancearse entre ambos grupos. Más adelante veremos que existen métodos capaces de remover el sesgo generado por las variables no observables.

¹⁰ En el caso del sesgo debido a desbalance en variables no observables, se puede hacer un análisis similar al presentado, por lo que no lo desarrollaremos aquí.

Los términos tercero y cuarto de la ecuación (\clubsuit) muestran el sesgo relacionado con el efecto directo de x sobre y . La suma de estos dos términos sería cero si $E(y | d, x = 0) = E(y | d, x = 1)$, lo que se interpreta como si la variable x no tuviera efecto sobre y . Asimismo este efecto también desaparece si x y d son estadísticamente independiente.

Ilustraremos estos sesgos mediante un ejemplo. Supongamos que se aplica un programa de concientización sobre los peligros del tabaco a una población de fumadores, donde la única característica observable a considerar es el sexo, siendo $x = 0$ para las mujeres y $x = 1$ para los hombres. Sea y el número de cigarrillos fumados por semana, y d indica como siempre la participación o no en el programa. Supongamos que en esta población $\Pr(x = 0) = 0,55$ y $\Pr(x = 1) = 0,45$, pero que en lo concerniente a la participación en el programa hay un desbalance por sexo, siendo $\Pr(x = 0 | d = 0) = 0,8$ y $\Pr(x = 1 | d = 1) = 0,7$. También:

$$E(y | d = 1, x = 0) = 21 \quad E(y | d = 0, x = 0) = 24$$

$$E(y | d = 1, x = 1) = 28 \quad E(y | d = 0, x = 1) = 30$$

De donde se puede ver que el efecto del programa para las mujeres es de $ATE|_{x=0} = -3$ cigarrillos menos por semana. Para los hombres, el efecto del programa es $ATE|_{x=1} = -2$ cigarrillos por semana. El ATE para toda la población es $-2,5$. Sin embargo, si reemplazamos los valores en la ecuación (\clubsuit) encontramos que:

$$E(y | d = 1) - E(y | d = 0) = 0,7$$

Es decir, la mera diferencia de medias del consumo promedio de cigarrillos entre los que siguieron el programa y quienes no lo siguieron arroja un sorprendente aumento en el consumo promedio de cigarrillos, pese a que por separado el consumo disminuyó tanto para hombres como para las mujeres. A este resultado se le conoce como la Paradoja de Simpson y es el resultado de la influencia del desbalance de x .

Por último, un estimador consistente de ATE en el caso de selección en observables (es decir, bajo el supuesto II') puede obtenerse mediante regresiones lineales. Tal como demuestra Wooldridge (2001), en la regresión donde x es un vector de variables, β_2 y β_3 son vectores de parámetros,

$$y_i = \beta_0 + \beta_d d_i + \beta_2' x + \beta_3' (d_i \cdot (x - \bar{x})) + u_i$$

El estimador $\hat{\beta}_d$ es un estimador consistente de ATE .

3. ESTUDIOS EXPERIMENTALES Y NO EXPERIMENTALES

Habiendo observado la importancia del cumplimiento de los supuestos identificadores de *ATE* y *ATET*, cabe la pregunta ¿bajo qué condiciones los datos que se utilizan para evaluar el impacto de políticas cumplirían estos supuestos? La respuesta a esta interrogante se halla en la forma como se generaron estos datos.¹¹

Como se ha visto, la aleatorización del tratamiento *d* hace que la diferencia de promedios sea un estimador consistente de *ATE* (y *ATET*). Pero, ¿de dónde proviene la idea de la aleatorización del tratamiento? Existe en la ciencia un procedimiento conocido como experimento aleatorio controlado, el cual se considera como el *gold standard* de la evaluación de impacto por cumplir (casi) perfectamente la condición de aleatorización de *d*.

3.1. EXPERIMENTOS ALEATORIOS CONTROLADOS

Estos experimentos tienen su origen en las ciencias biológicas y médicas, el cual consiste en el estudio del efecto de un tratamiento sobre un resultado de interés. Luego de haberse definido a una muestra aleatoria de individuos a ser estudiados, el procedimiento consiste en la selección aleatoria de dos subgrupos de individuos llamados «grupo tratamiento»¹² y «grupo de control». Al primero de ellos se les aplica intencionalmente el tratamiento del estudio, mientras que al segundo no recibe el tratamiento. Cuando el experimento es aplicado a seres humanos, usualmente al grupo de control se les entrega un placebo absolutamente inofensivo¹³ (por ejemplo, una píldora de similar característica a la recibida por los tratados), con el fin de evitar cualquier desviación en la conducta tanto de los tratados como de los no tratado. En este procedimiento, el tratamiento es independiente —estadísticamente hablando— de los resultados potenciales y de las características observables y no observables de los individuos, los cuales deberían estar estadísticamente balanceados entre ambos grupos. Por tal razón, los resultados del grupo no tratado simulan bien el escenario contrafactual en donde los tratados no reciben el tratamiento. Tal propiedad le da *validez interna* al estudio pues el resultado de la evaluación de impacto estaría libre de sesgos. Asimismo, si la muestra de individuos del análisis fue obtenida de manera aleatoria de la población de interés, asumiendo que no hay efectos de equilibrio general (supuesto SUTVA), estos resultados son generalizables a toda la población lo que le da *validez externa* al estudio.¹⁴

¹¹ Una interesante reseña de los problemas que se pueden encontrar en estudios observacionales —en comparación con estudios experimentales— se encuentra en el clásico documento de Cochran (1965).

¹² En las aplicaciones a la economía hemos llamado «grupo beneficiario» al grupo tratamiento.

¹³ El placebo podría no ser necesario si se trata de experimentos en animales o plantas.

¹⁴ Debe tenerse en cuenta que hay dos etapas de aleatorización: la primera ocurre en la selección de la muestra a ser sujeta de estudio de la población, y la segunda ocurre cuando el tratamiento es asignado aleatoriamente a un subgrupo de la muestra.

De ahora en adelante, usaremos el término «grupo de control» o grupo *C* a aquel conjunto de individuos no tratados que pueden representar bien el escenario contrafactual sin tratamiento. En el caso de los experimentos aleatorios controlados, los grupos de no tratados (grupo *N*) y control (grupo *C*) son exactamente iguales.

En lo que se refiere a la aplicación de estos experimentos a economía, dado que se estudia el impacto aislado de una variable sobre otra, este procedimiento es ideal para estudiar el efecto causal entre estas variables. Por ello no necesita el desarrollo de modelos teóricos que modelen la conducta de los agentes en el mundo real¹⁵, al ser sus implicancias absolutamente parciales¹⁶ y reducidas a dos variables. Esta alta precisión en la medición del impacto es a su vez una desventaja si el objetivo es tener una idea más completa del comportamiento de los agente, es decir, de los determinantes de los resultados observados.

La aplicación de este procedimiento resulta muy atractiva. No obstante para la evaluación de programas surgen algunos inconvenientes reportados en la literatura¹⁷. Acerca de la validez interna del procedimiento, usualmente es difícil encontrar un equivalente al placebo utilizado en medicina, por lo que en algunos casos es casi inevitable que los sujetos no solo noten que están siendo sujetos al experimento sino que noten a qué grupo pertenecen (*B* o *C*). Por ejemplo, si se tratara de un programa de capacitación laboral, sería muy extraño que se otorguen charlas de capacitación completamente inútiles a los miembros del grupo *C* con el fin que no noten que son beneficiarios. Este hecho puede provocar algunas problemas como la autoselección en los programas (pues es difícil que se pueda obligar a las personas a aceptar un tratamiento, el cual es normalmente voluntario) y al desgaste ocurrido (no al azar) por el abandono de algunos individuos a seguir en el programa. Existen otros problemas de orden ético si se trata de programas que podrían tener consecuencias en el largo plazo. Por ejemplo, si se pretendiera aplicar este procedimiento a programas que otorgan crédito educativo, para aquellas personas que no lo reciban podrían tener consecuencias negativas muy grandes por el resto de sus vidas (debido a la pérdida de oportunidades). Algo similar podría ocurrir con programas alimentarios. Todo esto genera serios cuestionamientos de orden ético para la aplicación de experimentos de este tipo para evaluar programas.

Pese a estas razones, en los últimos años se ha aplicado esta técnica en estudios de evaluación de impacto, principalmente de políticas económicas. Algunos ejemplos son: Gertler (2004), quien analiza el impacto del programa mexicano de transferencias condicionales de dinero —conocido entonces como PROGRESA— en la salud de los niños. En la implementación de este programa se seleccionó a 505 villas de zonas pobres de

¹⁵ Esto no significa de ninguna manera que se ignore a los modelos económicos y a los modelos probabilísticos subyacentes, quienes darán luz sobre el efecto que se espera y su explicación.

¹⁶ Entiéndase el término «parcial» en el sentido utilizado en economía en el análisis de estática comparativa.

¹⁷ Stock y Watson (2003) en su capítulo 11 cita algunos inconvenientes que se pueden encontrar en la práctica.

México, en donde en una primera etapa se escogió al azar a 320 poblaciones como beneficiarias y 185 como control. Esta forma de selección y el hecho de que el grupo de control no fue informado que sería en el futuro también beneficiario del programa le da características al experimento de ser muy parecido a un experimento aleatorio controlado. Con ello se cumple que estadísticamente hablando los grupos de beneficiarios y no beneficiarios se encuentren balanceados en sus características observables y no observables, lo que convence al autor para estimar el efecto causal mediante una regresión con una variable *dummy*, indicando la participación o no en el programa. Gertler menciona que además controla en la regresión por características socioeconómicas con el fin de mejorar el poder de las estimaciones y reducir la variación idiosincrática en la población.

En un estudio similar, Hoddinott y Skoufias (2003) también estudian el mismo programa pero esta vez para calcular el impacto sobre el consumo de alimentos. A diferencia de Gertler, estos autores son cautelosos con respecto a la aplicación inmediata de la diferencia de medias como estimador del efecto causal, al encontrar problemas de no cumplimiento en la muestra de beneficiarios (muchos hogares de localidades beneficiarias no recibieron tratamiento) y al haber significativas diferencias en cuanto a composición por género, tamaño del hogar y edad de los participantes entre los grupos de beneficiarios y no beneficiarios. Sugieren que los métodos de regresión controlados por estas variables son una mejor alternativa a la simple e incondicional diferencia de medias.

En otro estudio, como el de Angrist y Lavy (2002), se utiliza la extensión de esta metodología cuando la aleatorización se hace a niveles de grupos¹⁸ y no de individuos, existiendo entonces dos niveles de aleatorización. Esta metodología es usada ampliamente en medicina y psicología, y se aplica especialmente cuando se selecciona al azar a grupos (como, por ejemplo, comunidades, hospitales, escuelas, etc.), y luego en una segunda etapa se selecciona a individuos dentro de cada grupo. Esta metodología ha demostrado ser menos costosa en su implementación pero presenta menor poder estadístico que los muestreos aleatorios simples, principalmente debido a la correlación entre los grupos, entre otras deficiencias¹⁹. En el estudio de estos autores se analiza el efecto de un programa de premios monetarios a estudiantes sobre el rendimiento en una prueba académica específica (el Bagrut) en Israel y para ello conducen dos experimentos, uno donde la aleatorización se hace a nivel de individuos y otro en donde se realiza a nivel de escuelas. Cabe resaltar en este trabajo que los autores tuvieron inconvenientes en el momento de implementar la aleatorización a nivel de individuos debido a preocupaciones de los directivos educativos sobre la elección de los miembros del grupo beneficiario. Debido a esto se utilizó un mecanismo que además de utilizar al azar se tomaba en cuenta el estatus socioeconómico de los estudiantes. También existieron problemas para

¹⁸ Conocido en inglés como *Group Randomized Trials*. Apuntes sobre la metodología se pueden encontrar en Donner, Brown y Basher (1990).

¹⁹ La aleatorización de los grupos no garantiza el balanceo de las características a nivel de individuos, como sí lo hacen las aleatorizaciones de individuos. Para más observaciones, véase por ejemplo, Donner y Klar (2004).

implementar el experimento a nivel de escuelas, debiendo suspenderse en su primer año de aplicación debido a serias controversias desatadas en los medios y la opinión pública. Este es un claro ejemplo de las dificultades que este tipo de estudios experimentales debe enfrentar en el momento de su implementación práctica.

Existen numerosos ejemplos que utilizan datos experimentales, de los cuales solo mencionaremos a algunos por razones de espacio. Por ejemplo, Banerjee et al. (2004) realizan dos estudios aleatorizados evaluando el impacto de programas de asistencia educativa a estudiantes con bajo rendimiento sobre el aprendizaje medido como puntaje promedio de sus pruebas académicas. En otro trabajo, Angrist et al. (2002) estudian el efecto de la entrega de cupones a estudiantes secundarios sobre el rendimiento y la asistencia escolar en Colombia. Aunque este estudio no podría catalogarse estrictamente hablando como un experimento controlado, existió en él una asignación aleatoria (mediante un sorteo) de los cupones, lo que le dio el carácter de experimento «natural» y facilitó el cálculo del impacto. En la siguiente sección se explica en qué consisten estos experimentos naturales y los cuasiexperimentos.

3.2. ESTUDIOS OBSERVACIONALES

Lo que hace que un experimento sea «verdadero» y tenga el poder de medir correctamente el impacto de una variable sobre otra es la aleatorización en la selección de la muestra del estudio y en la aleatorización del tratamiento, ambos bajo el control del investigador. Como ya se explicó anteriormente, esta característica garantiza que los grupos de beneficiarios y no beneficiarios sean comparables. Como hemos visto en la sección anterior, en muchas ocasiones es difícil garantizar que el tratamiento se asigne en forma aleatoria de la forma como lo planeó el investigador. Normalmente suceden inconvenientes que afectan la validez interna del estudio. En otras ocasiones, por cuestiones prácticas es imposible asignar el tratamiento en forma aleatoria. Cada vez que tengamos un estudio en donde el tratamiento ha sido asignado en forma no aleatoria sino que se basa en observaciones fuera del control del investigador tendremos un estudio observacional.

Los estudios observacionales no son en sí mismos experimentos, aunque de alguna manera pretenden simularlos en el sentido que buscan elucidar una relación causal entre dos variables. En estos casos, si bien es posible distinguir una variable de tratamiento y una o más variables de resultados como posible consecuencia, tal tratamiento no ha sido asignado bajo el control del investigador. Por ejemplo, el tratamiento pudo ser resultado de cambios en la legislación que afectó a cierto sector de la población pero no a otro, a aspectos administrativos o quizás a cuestiones puramente naturales (fenómenos atmosféricos, telúricos, etc.).

En ocasiones, el tratamiento puede haber sido asignado en una forma no sistemática que se asemeja bastante bien a lo que hubiera sido un experimento controlado. En tal

caso se suele hablar de un experimento «natural». En cambio si el tratamiento está lejos de haber sido asignado en forma aleatoria pero el estudio realiza un importante esfuerzo por asegurar la comparabilidad de los tratados versus los no tratados, entonces tenemos un «cuasiexperimento»²⁰.

Una característica frecuente de los cuasiexperimentos (aunque no necesariamente indispensable para su definición) es que los grupos de beneficiarios y control ya existen como grupos definidos *antes* del tratamiento.

El hecho de que los datos se basen en observaciones genera un potencial problema de validez externa del procedimiento pues no existe la seguridad de que tales datos representen a la población total. Por ejemplo, los datos provenientes de programas de capacitación para el empleo podrían no representar a la población total de desempleados si la evaluación del programa se concentra en determinadas áreas geográficas (grandes ciudades, por ejemplo), o si cuentan con medios de información para estar al tanto del programa.

Asimismo, una asignación del tratamiento fuera del control del investigador presenta un potencial problema de validez interna si es que este tratamiento no es asignado en forma aleatoria²¹. Como vimos en la sección anterior, esto representaría una violación al supuesto (I) —y por ende al (II)— de la sección 2.2, lo cual invalidaría el cálculo del *ATE* mediante la diferencia de medias pues no habría la garantía que los grupos *B* y *N* sean comparables. Es por ello que se requiere de un tratamiento estadístico muy cuidadoso con el fin de replicar o simular una situación de tratamiento aleatorio, o en otro caso, habiéndose entendido las razones para la no aleatorización, tomarlas en cuenta con el fin de obtener estimaciones válidas.²²

A pesar de estas dificultades, existen numerosos estudios que utilizan datos cuasiexperimentales que buscan replicar los resultados de los experimentos controlados, debido fundamentalmente a las ventajas que estos estudios tienen en términos de acceso a datos y a que podrían no sufrir algunos de los efectos perversos que contaminan a los experimentos aleatorios controlados²³. Estos estudios son de diferente naturaleza, algunos en si mismos pueden ser parecidos a los experimentos aleatorios controlados, mientras que otros pueden ser bastante distintos. En las siguientes secciones veremos estrategias que se adaptan a diversos problemas presentados a la hora de evaluar el impacto cuando se tiene este tipo de datos.

²⁰ Véase Rosenbaum (2009), 4-6. Es frecuente encontrar en la literatura que los términos «experimentos naturales» y «cuasiexperimentos» son usados como sinónimos.

²¹ Sin embargo no se descarta que en algunos casos excepcionales de cuasi experimentos, el tratamiento sí haya sido asignado en forma aleatoria por pura cuestión del azar.

²² Tal como menciona Campbell (1969), 412, sobre este problema: «The general ethic, here advocated for public administrators as well as social scientists, is to use the very best method possible, aiming at 'true experiments' with random control groups. But where randomized treatments are not possible, a self-critical use of quasi-experimental designs is advocated. We must do the best we can with what is available to us.»

²³ Por ejemplo la llamada «reactividad» que se refiere al cambio en la conducta de las personas sujetas al estudio como el *Hawthorne Effect*.

4. DISEÑO DE REGRESIÓN DISCONTINUA AGUDA

En un destacado estudio, Angrist y Lavy (1999) analizaron una manera cómo identificar el efecto del número de estudiantes por aula en escuelas de Israel sobre el rendimiento educativo utilizando datos observacionales. Los autores estudiaron la «regla de los Maimonides», en la cual ninguna escuela de Israel debería tener aulas con más de 40 alumnos. Si la matrícula excede ese número, inmediatamente se divide el aula en dos secciones con algo más de 20 alumnos por aula. En ese sentido, la regla estaría seleccionando (casi al azar) a un grupo de estudiantes a estudiar en aulas cercanas (por la izquierda) a los 40 alumnos y a otro grupo de similares características a estudiar en clases de menor tamaño. Asumiendo que la matrícula total no está relacionada con las características de los estudiantes, la única diferencia entre los dos grupos mencionados sería el tamaño del aula promedio. Con ello se lograría identificar el efecto tratamiento promedio al menos localmente alrededor de la discontinuidad en el tamaño de la clase.

Aunque el estudio original de Angrist y Lavy muestra que tal discontinuidad en la práctica no está tan claramente definida como lo señala la regla (lo cual requiere algunas correcciones adicionales que veremos más adelante), la importancia de esta nueva corriente estriba en que se puede identificar el efecto tratamiento promedio al menos localmente alrededor de la discontinuidad de una variable, siempre y cuando se cumplan algunas condiciones básicas. Estas son que las entidades se encuentren ordenadas en forma continua con respecto a una *variable índice* (en este caso la matrícula), y que la variable resultado (en este caso, el rendimiento escolar) también esté relacionada continuamente con la variable índice, además se observa una asignación del tratamiento con respecto a un umbral definido sobre la variable índice, lo cual genera una discontinuidad en el resultado observado en función del índice. Debido a la similitud de los individuos por encima o debajo del umbral, el salto en el resultado es el efecto tratamiento promedio alrededor del umbral. Esta es la base general de los diseños de regresión discontinua aguda.

Los diseños de regresión discontinua son un caso especial de experimentos naturales en donde es posible identificar el efecto promedio del tratamiento al menos localmente. En la literatura reciente de evaluación de programas se ha venido aplicando esta técnica de regresión discontinua, la cual ha sido desarrollada y sistematizada en décadas recientes por Hahn, Todd y van der Klaauw (2001) y otros²⁴.

Formalizando lo mencionado en los párrafos anteriores²⁵, en el contexto del modelo de resultados potenciales, supongamos que el vector de variables observables para cada entidad i se compone de (X_i, Z_i) , donde X_i es un escalar y Z_i es un vector de las demás características observables de i que se asume que no han sido afectadas por el tratamiento.

²⁴ Un par de documentos clásicos que muestran el uso de este enfoque en la década de 1960 son los de Thistlethwaite y Campbell (1960) y Campbell (1969), pero recién en años recientes este método ha recobrado popularidad. Pueden consultarse algunas referencias destacables como Imbens y Lemieux (2007).

²⁵ Aquí seguimos el desarrollo de Imbens y Lemieux, y el de Lee (2005).

A la variable X (la cual debe ser una variable continua) se le conoce como *forcing variable* y es la variable índice que se mencionó líneas arriba, pues los valores del tratamiento d_i se encuentran completamente determinados por los valores de X si se encuentran a un lado o al otro de un umbral fijo c .

La idea general es que dado este punto de corte, si la relación entre X y los resultados potenciales y_j es suave, cualquier discontinuidad observable en $E[y | X]$ será el efecto del tratamiento en el punto c .

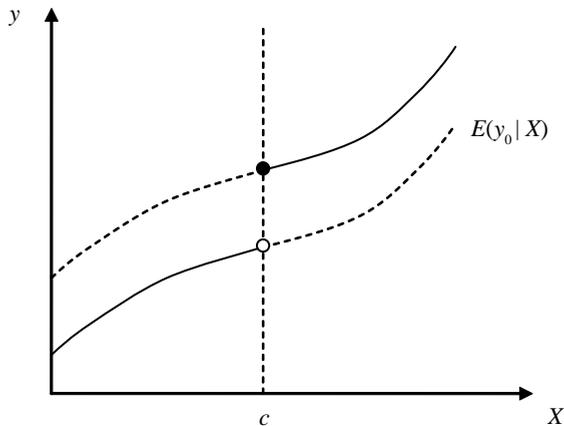
En este caso, ocurrirá que $d_i = 1[X_i \geq c]$ donde 1 es el operador que otorga el valor de 1 si es verdad la condición mencionada y 0 en otro caso, y c es un punto de corte definido exógenamente. Es frecuente que la variable X sea re-escalada con tal que el punto de corte se ubique en cero.

Evidentemente este es un caso extremo de selección en observables pues los grupos B y N difieren absolutamente en la variable X y, por lo tanto, la diferencia de las medias de grupo no es un estimador apropiado del efecto del tratamiento. Por el contrario, la idea del método es poder identificar el efecto tratamiento al menos localmente alrededor de c .

Gráficamente, las líneas punteadas indican la esperanza condicional de los resultados potenciales dado X , $E[y_j | X]$ para $j = 0, 1$. Mientras tanto, la línea continua indica la esperanza condicional del resultado observado, el cual matemáticamente es:

$$E[y | X] = E[y | d = 0, X] \cdot \Pr[d = 0 | X] + E[y | d = 1, X] \cdot \Pr[d = 1 | X]$$

Gráfico 1
Ilustración de una regresión discontinua aguda

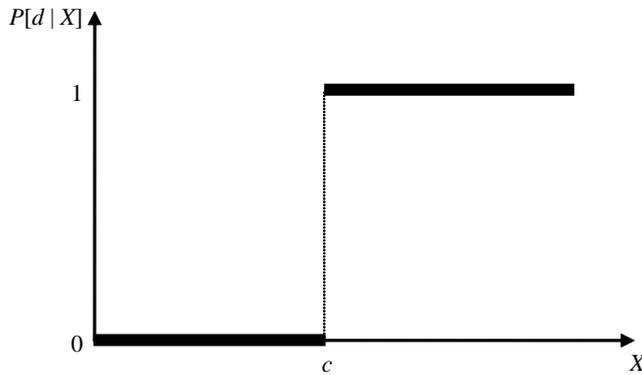


En el gráfico 1, el efecto tratamiento es el salto en la esperanza condicional de y dado X .

$$\lim_{x \downarrow c} E[y_i | X_i = x] - \lim_{x \uparrow c} E[y_i | X_i = x]$$

Es ilustrativo dibujar también la relación entre X y la probabilidad de recibir el tratamiento dado X , o sea $P[d | X]$. Esta relación se muestra en el gráfico 2.

Gráfico 2
Probabilidad de recibir el tratamiento



En este contexto de regresión discontinua aguda, seguimos asumiendo que se cumple (II') pero ya no se está cumpliendo el supuesto de *matching* o *overlapping*, $0 < P(d | X) < 1$.

La violación de este supuesto genera algunos inconvenientes en la identificación el efecto del tratamiento. El efecto tratamiento promedio en el punto c es:

$$ATE = E[y_1 - y_0 | X = c] = E[y_1 | X = c] - E[y_0 | X = c]$$

El primer término es estimable con cierta dificultad pues se requiere que exista un número significativo de observaciones de $X = c$, lo cual puede no cumplirse pues X es continua. En el caso del segundo término, no hay datos de y_0 para $X = c$ por definición. Se hace entonces imperioso que la identificación de estos efectos se haga localmente alrededor de c .

Es por ello que se hacen dos supuestos de regresión discontinua aguda que permiten identificar el efecto (adicionales al supuesto (II')):

Continuidad de la función de regresión condicional: $E[y_j | X = x]$ es continua en x , para $j = 0, 1$.

Continuidad de la función de distribución condicional: Sea $F_{y_j | X}(a | b) = \Pr(y_j < a | X = b)$, asumimos que $F_{y_j | X}(a | b)$ es continua en a para todo b , para $j = 0, 1$.

Bajo estos supuestos:

$$E[y_0 | X = c] = \lim_{X \uparrow c} E[y_0 | X] = \lim_{X \uparrow c} E[y_0 | d = 0, X] = \lim_{X \uparrow c} E[y | X]$$

y similarmente:

$$E[y_1 | X = c] = \lim_{X \downarrow c} E[y_1 | X] = \lim_{X \downarrow c} E[y_1 | d = 1, X] = \lim_{X \downarrow c} E[y | X]$$

Luego, el ATE es:

$$ATE = \lim_{X \downarrow c} E[y | X] - \lim_{X \uparrow c} E[y | X]$$

El hecho de afirmar que el efecto tratamiento es identificable localmente alrededor de c indica que estamos suponiendo que los individuos alrededor de c son comparables, tanto en sus características observables como no observables (balanceo en ambas características). Es decir, es como si el tratamiento se hubiera asignado aleatoriamente alrededor de c , lo cual le da el carácter de experimento natural a este diseño. A este supuesto también se le conoce como el supuesto de «intercambiabilidad». Este supuesto podría no cumplirse (lo cual invalidaría la identificación del *ATE*) si los individuos pudieran alterar su información observable de X con el fin de recibir o no el tratamiento. Por ejemplo, si se tratara de un programa de ayuda para individuos con ciertos ingresos por debajo de un umbral c . Si aquellos individuos que estuvieran apenas por encima de c alteraran su ingreso reportado reduciéndolo para hacerse beneficiarios del programa, ocurriría entonces que por debajo de c se agruparían individuos con ciertas características no observables, quedando por encima de c a otros individuos que difícilmente serían comparables con los «de abajo».

Finalmente, el *ATE* se puede estimar mediante una regresión semi-lineal del tipo:

$$y_i = \beta_d d_i + g(X_i) + u_i$$

Donde $g(X_i)$ es una función continua en $X = c$. Aquí se cumple que:

$$ATE = \lim_{X \downarrow c} E[y | X] - \lim_{X \uparrow c} E[y | X] = \beta_d$$

En resumen, hemos mostrado cómo es posible identificar el efecto tratamiento en el caso de la regresión discontinua aguda. No obstante, los supuestos de este modelo podrían no cumplirse en la realidad especialmente cuando se trata del cumplimiento de la regla de asignación del tratamiento. Es decir, es frecuente que el umbral que define al tratamiento no sea tan claro como se menciona aquí sino que algunas entidades no cumplan con la regla, existiendo cierta probabilidad de recibir el tratamiento estando por debajo del umbral, o de no recibirlo estando por encima de él. El desarrollo de este caso se presenta en la sección 6.

Existen algunos ejemplos de la aplicación de regresión discontinua aguda en estudios de impacto. En la literatura de elecciones, se ha encontrado que en elecciones ajustadas (donde las dos opciones a elegir se encuentran alrededor del 50%) se crea una discontinuidad aguda (o determinística) que puede ser explotada para la identificación local del impacto. Ejemplos de esta literatura son Lee, Moretti y Butler (2004), quienes estudian la conducta (como congresistas) de diputados del Partido Demócrata de los Estados Unidos que han sido elegidos en elecciones ajustadas, en contraste con diputados del Partido Republicano que también han sido elegidos en elecciones con escasa diferencia. Asumiendo que en este tipo de elecciones ajustadas, una variabilidad natural hace que a veces gane uno u otro candidato (lo cual mostraría que la fuerza de cada partido en su distrito es más o menos similar), se observa que las votaciones en el Congreso de esos diputados elegidos no se acerca a la media, sino que ellos adoptan posiciones ya sea un

tanto más de izquierda para los demócratas y de derecha para los republicanos (esto último medido mediante un índice). Es decir, encuentran una significativa discontinuidad en el récord de votaciones de estos diputados, lo cual muestra que al ser elegidos los políticos no representan la voluntad de la población que los eligió sino que siguen sus propios designios ideológicos y los de su partido.

En otro trabajo que aplica esta metodología, DiNardo y Lee (2002) estudian el efecto de la sindicalización de empresas sobre la supervivencia de las mismas. Los autores explotan el hecho de que las empresas se sindicalizan si consiguen al menos el 50% más 1 de los votos de los trabajadores. Con ello, muestran con datos que las empresas con porcentajes de votación a favor de la sindicalización alrededor del 50% tienen características observables muy similares, con lo cual se puede atribuir el hecho de estar sindicalizado o no a factores aleatorios. Encuentran que el efecto de la sindicalización sobre la supervivencia de las empresas es muy pequeño.

En otro tipo de estudios económico-sociales, Barrera-Osorio, Linden y Urquiola (2007) encuentran que el programa colombiano *Gratuidad*, el cual libera del pago de derechos académicos a estudiantes según ciertos niveles socioeconómicos pre-establecidos, presenta datos que pueden ser estudiados mediante regresiones discontinuas. La discontinuidad se ubica en los niveles socioeconómicos discretos (nivel I, nivel II, etc.) definidos sobre un índice de pobreza continuo llamado *Sisben*. En tal sentido, individuos alrededor de los cortes (por ejemplo, alrededor de los 11 puntos de *Sisben*) pueden ser considerados muy similares, sin embargo, aquellos a la izquierda de 11 se benefician del programa, pero aquellos a la derecha de 11 puntos no reciben el beneficio. Se considera entonces que la única diferencia entre este grupo alrededor de los 11 puntos es el programa, lo cual identifica el efecto causal. Encuentran que participar en el programa incrementa en 3% la probabilidad de matricularse en la escuela para el grupo de estudiantes de educación básica.

5. EL MÉTODO DE PAREO O *MATCHING*

El método de pareo es una técnica muy popular usada en el análisis de políticas a través de datos no experimentales. A diferencia del diseño de regresión discontinua aguda en donde los grupos de tratados y no tratados se encuentran completamente separados según la variable índice, en este caso se cumple el supuesto *overlapping* o *matching* (ver sección 2.4) y por ello los individuos de los grupos *B* y *N* comparten ciertas características en un rango común.

Si bien es cierto que es una técnica estadística relativamente antigua, en años recientes ha tenido importantes avances y perfeccionamientos en econometría (por ejemplo, véase Heckman, Ichimura y Todd 1997, 1998). En términos generales, busca evitar el problema del *confounder* en estudios con datos observacionales (véase sección 2.5) que ocurre cuando el efecto del tratamiento sobre el resultado no puede ser distinguido del

efecto de una tercera variable relacionada con las dos primeras, debido al desbalance de esta variable en los grupos B y N . Para lograr ese objetivo, el método del pareo, mediante la conformación de parejas, busca definir un subgrupo de no beneficiarios (grupo de control C) tal que cualquier variable confundidora quede balanceada entre los tratados y los controles. Sin embargo, el método solo logra evitar el sesgo generado por variables confundidoras observables.

Esta técnica es especialmente útil cuando:

- (1) Se busca estimar el $ATET$
- (2) Se posee un número grande de individuos en el conjunto N
- (3) Se posee un conjunto rico de variables observables, en especial antes de la aplicación del tratamiento

Observando el $ATET = E(y_1 - y_0 | d = 1) = E(y_1 | d = 1) - E(y_0 | d = 1)$. Como se mencionó antes, el término $E(y_0 | d = 1)$ no es observable. Además si el tratamiento no ha sido asignado en forma aleatoria —como suele ser el caso en los cuasiexperimentos— no podemos utilizar a un estimador de $E(y_0 | d = 0)$ como una aproximación de $E(y_0 | d = 1)$, pues nada garantiza que las características observables y no observables se encuentren balanceadas entre los grupos de beneficiarios y no beneficiarios.

Ante este problema de identificación, el método propone unos supuestos identificadores, bajo los cuales sería posible calcular el $ATET$. Empecemos asumiendo los supuestos llamados de *strong ignorability* por Rosenbaum y Rubin (1983) (I'): $(y_0, y_1) \perp\!\!\!\perp d | x$, y (III'): $0 < \Pr(d | x) < 1$, asumiéndose que estas propiedades se cumple en algún subconjunto \mathcal{X} de valores del dominio de x , el cual llamaremos el soporte común²⁶. En términos intuitivos, el primer supuesto quiere decir que si controlamos a los individuos según sus características observables (por ejemplo, su género), en cada subgrupo que corresponde a valores específicos de x el tratamiento es independiente de los resultados, es decir ha sido asignado de forma similar a una asignación aleatoria. El segundo supuesto afirma que, para cada valor de características observables x , existen individuos que han sido tratados y otros que no han recibido el tratamiento.

Bajo estos supuestos, condicionado a x podemos utilizar al grupo no beneficiario N como el escenario contrafactual buscado (grupo de control, C). Entonces, si se cumple el supuesto (I') para $x \in \mathcal{X}$, tendremos que

$$E(y_0 | x, d = 1) = E(y_0 | x, d = 0)$$

Si adicionalmente se asume (II) en $x \in \mathcal{X}$, tendremos que (Lechner 2001):

$$E(y_0 | d = 1) = E_x(E(y_0 | x, d = 0) | d = 1)$$

²⁶ De hecho, para identificar a $ATET$ basta con asumir que $y_0 \perp\!\!\!\perp d | x$.

Que puede ser estimado mediante análogos muestrales. No obstante, tal como Heckman, Ichimura y Todd (1997, 1998) han remarcado, los supuestos de *strong ignorability* son innecesariamente fuertes, pues basta con el supuesto de independencia de medias condicionales (II') para la identificación de *ATE*.

Cabe mencionar que si hay características no observables de los individuos que no están balanceadas, entonces la diferencia de medias condicionada a x no sería un buen estimador del efecto tratamiento promedio. Esta suele ser la principal deficiencia de esta técnica.

Pero asumiendo que se cumplen las condiciones de *strong ignorability*, la discusión restante es cómo encontrar dentro de los no tratados a un grupo de control que comparta las mismas características que el grupo beneficiado y que pueda ser utilizado como el escenario contrafactual.

5.1. PAREO EXACTO E INEXACTO

La respuesta a la interrogante planteada nos lleva a preguntarnos si efectivamente existirán individuos que tengan las mismas características pero que pertenezcan a grupos distintos según la recepción del tratamiento.

Una primera forma de hacer esta búsqueda es mediante el pareo exacto. Para cada unidad $i \in B$ con características x_i , se busca una unidad $j \in N$ que posea las mismas características, es decir $x_i = x_j$. Los «pares» de cada unidad i tomando como base a las características x son aquel grupo $A_i(x) = \{j \in N \mid x_i = x_j\}$. Luego el grupo de control es la unión de todos los conjuntos A_i , es decir $C = \bigcup_{i \in B} A_i$.

Esta forma de hacer pareo tiene un problema conocido como el «problema de la dimensionalidad». Puesto que en los estudios con datos microeconómicos los individuos suelen tener muchas características observables, es posible que para muchas unidades i no exista su par exacto j que comparta todas esas características (por ejemplo, la edad, el sexo, el nivel educativo, etc.) y, por lo tanto, el grupo de control C podría tener muy pocos elementos o quizás ninguno.

Una alternativa a la versión exacta del pareo es la llamada «inexacta», en donde se busca a unidades que sean parecidas a las tratadas, aunque no lleguen a tener exactamente las mismas características. Para ello se definen unos criterios de cercanía. En este contexto, los «pares» de la unidad i son el grupo $A_i(x) = \{j \in N \mid x_j \in v(x_i)\}$ donde $v(x_i)$ define a una vecindad cercana a x_i .

Las unidades cercanas a i podrían ser numerosas, por ello se suele simular al escenario contrafactual y_0 con el promedio de estas unidades cercanas²⁷. Para realizar este cálculo se acostumbra promediarlos usando ponderadores $\omega(i, j)$ con $0 \leq \omega(i, j) \leq 1$

²⁷ Tal como señalan Dehejia y Wahba (2002), el hecho de tener un grupo de comparación unitario o numeroso no es un asunto trivial. Tener muchas unidades de comparación incrementa la precisión de la estimación del escenario contrafactual pero genera sesgos debido a que se utilizan unidades que podrían ser muy diferentes a la unidad tratada.

y $\sum_{j \in A_i} \omega(i, j) = 1$. Normalmente los ponderadores estarán relacionados con la cercanía de j a i , dándole mayor peso a los que se encuentren más cerca.

Es importante notar en este momento que el pareo podría hacerse con reemplazo o sin reemplazo. Si se hace sin reemplazo, una unidad j no beneficiaria no puede ser utilizada para reconstruir el escenario contrafactual de dos unidades $i \in B$ distintas. Puesto que esto trae problemas con pérdidas de observaciones, el pareo con reemplazo sí permite que una unidad j pueda ser utilizada más de una vez, siendo de especial utilidad cuando se tienen pocas observaciones (Dehejia y Wahba 2002).

En cualquier caso, la fórmula general del estimador de *ATE*T con pareo inexacto es:

$$\widehat{ATE}T = \frac{1}{n_B} \sum_{i \in B} \left(y_{1i} - \sum_{j \in A_i} \omega(i, j) y_{0j} \right)$$

Veamos a continuación algunos casos especiales de pareo inexacto, los cuales difieren ya sea en la conformación del grupo A_i a través de la definición de vecindad, o difieren en los pesos asignados en $\omega(i, j)$.

Un caso muy común de pareo es aquel que se realiza según el «vecino más cercano» (*nearest neighbor*). Se escoge a la unidad j que está más cerca de i usando la distancia euclídea. En este caso:

$$A_i(x) = \left\{ j \mid \min_j \|x_i - x_j\| \right\}$$

Normalmente este conjunto debería tener solamente un elemento ($\omega(i, j) = 1$ para el j más cercano y $\omega(i, j) = 0$ para cualquier otra unidad), aunque podría tener a más de uno si se produce un empate. Asimismo el investigador puede definir una distancia mínima (llamada «caliper») como primer filtro, con el fin de hacer un pareo con individuos que estén realmente cercanos.

5.2. PAREO MEDIANTE EL *PROPENSITY SCORE*

Una forma alternativa de resolver el problema de la dimensionalidad es creando un puntaje o *propensity score* que resuma en una sola variable a todas las características x de los individuos. En términos más específicos, el *propensity score* es la estimación de la probabilidad de ser beneficiario del programa, $P(x) = \Pr(d = 1 \mid x)$. En un estudio muy celebrado, Rosenbaum y Rubin (1983) demostraron que si bajo los supuestos de *strong ignorability* en el soporte común, entonces se cumple que $(y_1, y_0) \perp d \mid P(x), \forall x \in \mathcal{X}$. Este resultado es de mucha importancia pues permite que el pareo se pueda hacer basándose en el *propensity score* (Dehejia y Wahba 1999, 2002).²⁸

²⁸ Nuevamente, Heckman, Ichimura y Todd (1997) demuestran que $E(y_0 \mid P(x), d = 1) = E(y_0 \mid P(x), d = 0)$ es suficiente para identificar a *ATE*T.

En la práctica, el *propensity score* es estimado mediante regresiones *logit* o *probit*. Una vez hecha esta estimación, se puede hacer un pareo mediante, por ejemplo, el vecino más cercano en términos de este puntaje. En este caso, tendríamos que el conjunto de unidades «pares» a una unidad beneficiaria i es:

$$A_i(P(x)) = \left\{ j \in N \mid \min \left\| \hat{P}_i(x) - \hat{P}_j(x) \right\| \right\}$$

Normalmente este conjunto será unitario pues el *propensity score* es una variable continua que cuenta con un número ilimitado de decimales. Al igual que antes es posible definir una distancia mínima, $\|\hat{P}_i(x) - \hat{P}_j(x)\| < \tau$, pudiendo entonces ser el conjunto $A_i(x)$ vacío.

Una alternativa es la conocida como *radius matching*, en donde:

$$A_i(P(x)) = \left\{ j \in N \mid \|P_i - P_j\| < r \right\}$$

A diferencia del vecino más cercano, en el caso de *radius matching*, el conjunto $A_i(P(x))$ puede tener más de un elemento. El *ATET* se estima considerando el promedio simple de los resultados y de los elementos de $A_i(P(x))$.

Un problema con los métodos del vecino más cercano y *radius matching* es que consumen mucha información y pierden muchas observaciones, las cuales podrían contener información valiosa en la estimación de los escenarios contrafactuales. Una alternativa propuesta en la literatura es que se permita que las unidades del grupo de comparación $A_i(P(x))$ sean muchas alrededor del valor de x , pero ponderándolas según una función ponderadora llamada kernel²⁹, que da más peso a unidades cercanas y menor peso a las alejadas³⁰. Luego el ponderador $\omega(i, j)$ es:

$$\omega(i, j) = \frac{k\left(\frac{P_i - P_j}{h}\right)}{\sum_{j \in N} k\left(\frac{P_j - P_i}{h}\right)}$$

Donde P es el *propensity score*, $k(\cdot)$ es un kernel³¹ y h es el ancho de la «ventana» el cual determina cuantos valores P_j alrededor de P_i serán incluidas en el cálculo del promedio, es decir h define implícitamente a una vecindad. Esta especificación significa que el escenario contrafactual es estimado a través de la estimación de la esperanza condicional

²⁹ Un kernel es una función $k(x)$ que cumple algunas propiedades específicas. (i) $k(x)$ es simétrica alrededor de 0 y continua; (ii) $\int k(z) dz = 1$, $\int zk(z) dz = 0$, $\int |k(z)| dz < \infty$; (iii) $k(z) = 0$ si $|z| \geq z_0$ para un z_0 definido, o $|z|k(z) \rightarrow 0$ cuando $|z| \rightarrow \infty$; y (iv) $\int z^2 k(z) dz = k < \infty$.

³⁰ Véase Heckman, Ichimura, Smith y Todd (1998).

³¹ Algunos ejemplos de funciones kernel muy utilizadas son la uniforme donde $k(z) = (1/2) \cdot 1[|z| < 1]$, la triangular con $k(z) = (1 - |z|) \cdot 1[|z| < 1]$, la Epanechnikov donde $k(z) = (3/4) \cdot (1 - z^2) \cdot 1[|z| < 1]$, y la Gaussiana donde $k(z) = (2\pi)^{-1/2} \exp(-z^2/2)$.

de y sobre x mediante una regresión no paramétrica de y sobre x para las unidades del grupo no beneficiario. Esta regresión no paramétrica calcula el promedio simple de y en el intervalo seleccionado h . Una alternativa usada es la *regresión lineal local*, la cual calcula no solo un intercepto sino también una pendiente localmente en la vecindad.

En cualquiera de los dos casos, en la elección del ancho de la ventana h existirá un *trade-off* entre eficiencia y sesgo, pues una ventana más amplia abarca más observaciones lo cual genera una mayor eficiencia en las estimaciones, pero incrementa el sesgo que se origina en la estimación al suavizarse una curva³². En el extremo caso que $h \rightarrow \infty$, el valor de la regresión no paramétrica simplemente entregaría el promedio de los valores de y del grupo no beneficiario, lo cual estaría alejado de la media condicional de y dado x . Por otro lado, si h es muy pequeño, se contaría con muy pocas observaciones lo cual resta confiabilidad a las predicciones.

Cabe resaltar que tal como señalan Heckman, Ichimura, Smith y Todd (1998), una notable diferencia entre esta técnica en comparación con los experimentos aleatorios controlados es en el grupo de control que se genera. Mientras que en los experimentos, dada la naturaleza del proceso, se garantiza que las características observables y no observables tienen la misma distribución entre los beneficiarios y controles, en el caso de datos no experimentales nada garantiza que eso ocurra. Por ello es frecuente que el *propensity score* no tenga el mismo soporte que entre beneficiarios y no beneficiarios. Por tal razón, y con el fin de excluir a individuos que no tienen un «par» en el otro grupo, es necesario definir el rango o soporte común (*common support*) que hemos mencionado antes, que es la intersección de los soportes de los beneficiarios y no beneficiarios en sus *scores*. El pareo se va a realizar finalmente solamente entre aquellos individuos que tengan un *score* dentro de dicho rango común, eliminándose a todos los individuos que queden fuera de él.

Cuando se realiza un pareo de uno-a-uno (como en el caso del vecino más cercano) después de la definición del soporte común, la distribución del *score* entre los beneficiarios y el grupo de control debería ser muy parecida. Si el rango en común es muy pequeño o inexistente entre los *propensity scores* de los beneficiarios y no beneficiarios, no se podrá realizar el pareo y además será una señal clara que los dos grupos no son comparables.

Finalmente la expresión general del estimador del *ATET* con la definición del soporte común es:

$$AT\hat{E}T = \frac{1}{n_B} \sum_{i \in B \subset CS} \left(y_{1i} - \sum_{j \in A_i \subset CS} \omega(i, j) y_{0j} \right)$$

Donde *CS* hace referencia a que solo se toma en cuenta a individuos cuyas características pertenecen al soporte común.

³² En Caliendo y Kopeinig (2005) se pueden encontrar algunos consejos prácticos a tomar en cuenta para la implementación del *propensity score matching*.

Las aplicaciones del método de pareo en economía son numerosísimas y los campos en los que se utiliza crecen día a día. En economía laboral, en especial en lo que se refiere a programas de desempleo y entrenamiento, pueden consultarse por ejemplo los mencionados trabajos de Heckman, Ichimura y Todd (1997), Dehejia y Wahba (2002), Lechner (2000) y Burga (2003) en el caso del programa peruano Projovent.

Existen algunos temas adicionales acerca del Método de Pareo que no se presentarán aquí. Un tema crucial es la principal desventaja del método propuesto en su incapacidad para controlar el sesgo en variables no observables. Para evitar este inconveniente, la literatura se ha apoyado en el método tradicional de *diferencias en diferencias*, el cual desarrollaremos más adelante. Acerca de la elección entre los algoritmos propuestos para la construcción del escenario contrafactual, algunos estudios los han comparado encontrando algunas ventajas o desventajas entre ellos³³. Un resumen de estas comparaciones se puede encontrar en Vinha (2006). Sobre el tipo de tratamiento, es posible extender el procedimiento para el caso de tratamientos múltiples no binarios, en donde es importante la intensidad o heterogeneidad del mismo. Algunos estudios han desarrollado este análisis que aun es relativamente nuevo en economía (véase por ejemplo Joffe y Rosenbaum 1999, Imbens 2000, Lechner 2002).

6. ENDOGENEIDAD DEL TRATAMIENTO: EL MÉTODO DE VARIABLES INSTRUMENTALES³⁴

En ocasiones aun si los programas son diseñados para ser asignados en forma aleatoria entre la población objetivo, en la práctica la recepción o no del tratamiento está en manos de las personas quienes podrían decidir no recibirlo o logran recibirlo sin haber sido pre-seleccionados. En tal situación, las decisiones de las personas influyen en la variable tratamiento d , por lo cual se le debe considerar como una variable endógena. De esta manera no se cumplen los supuestos identificadores del *ATE*—debido a la autoselección generada— por lo que la diferencia de medias $\hat{\delta}$ ni la estimación por mínimos cuadrados ordinarios $\hat{\beta}_d$ en $y_i = \beta_0 + \beta_d d_i + u_i$ son estimadores consistentes del *ATE*.

La variable d podría no expresar plenamente el objetivo de la política pues algunas personas podrían decidir participar o no en ella; es decir, d dependería de algunas variables no observables de preferencias, las cuales están capturadas en el término de error u .

A manera de ejemplo (Angrist 1990; Angrist, Imbens y Rubin 1996), imaginemos que y es algún resultado (por ejemplo, el nivel educativo alcanzado por una persona i) y d indica si el individuo participó o no en el servicio militar. Esta variable d no ocurre al azar entre los individuos pues la participación depende de la decisión de ellos.

³³ Un resumen de estas comparaciones se puede encontrar en Vinha (2006).

³⁴ Una revisión muy intuitiva del método de variables instrumentales se encuentra en Angrist y Krueger (2001).

Supongamos que existe una variable binaria z relacionada con d pero que no con u . Por ejemplo, z podría representar un sorteo para designar a los elegidos para el servicio militar. No todos los sorteados hacen el servicio militar ni todos los que hacen el servicio fueron sorteados pero es claro que existe una asociación entre el sorteo y la participación en el servicio. Este no cumplimiento de lo que indica el sorteo puede deberse a muchas razones, como problemas de salud o motivación. El no cumplimiento de la intervención genera endogeneidad en la variable d . Es bastante claro que en aplicaciones a la economía, en especial en programas sociales, este problema del no cumplimiento es de suma importancia.³⁵

Luego, si z cumple las condiciones usuales de las variables instrumentales (está correlacionada con d pero no con u), entonces podemos identificar y estimar al parámetro β_d como el *ATE* en el modelo $y_i = \beta_0 + \beta_d d_i + u_i$. Calculando $\text{cov}(y, z) = \text{cov}(\beta_0 + \beta_d d + u, z) = \beta_d \text{cov}(d, z)$, tendremos que el parámetro poblacional $\beta_d = \text{cov}(y, z) / \text{cov}(d, z)$. Un análogo muestral de esta expresión es un estimador de variables instrumentales del efecto tratamiento promedio β_d .

El rol de z en la identificación del *ATE* mediante variables instrumentales se ubica en que extrae aquella variabilidad de d que no está relacionada con u , y la asocia con la variabilidad de y relacionada a z . Si bien es cierto que esta estrategia econométrica es útil, tiene un costo en términos de la pérdida de información que contienen las variables y y d .

Para estudiar este último aspecto, vale la pena preguntarse ¿qué es lo que logra identificar exactamente el estimador de variables instrumentales?³⁶ La respuesta a esta pregunta se basa en el análisis de la heterogeneidad de las respuestas de los individuos ante el instrumento. Siguiendo con el ejemplo del servicio militar, en la siguiente tabla se muestra los valores de d (en las filas) condicionados a valores de z (por columnas). Se distinguen cuatro tipos de individuos:

	$z = 1$	$z = 0$	Tipo
Valores de d	1	0	<i>Cumplidores</i>
	1	1	<i>Siempre tomadores</i>
	0	0	<i>Nunca tomadores</i>
	0	1	<i>Desafiantes</i>

³⁵ En la literatura médica se propone un análisis de la intención del tratamiento (*intention-to-treat analysis*) en donde se compara el resultado promedio de aquellos seleccionados por el programa versus el de aquellos no seleccionados por el programa, sin importar si cumplieron o no con el tratamiento.

³⁶ Un desarrollo más general se encuentra en Imbens y Angrist (1994). En esta parte seguiremos las exposiciones simplificadas que son estándares en la literatura (Lee 2005, Wooldridge 2001).

Los *cumplidores* hacen lo que dice el programa. Los *siempre tomadores* participan en el programa salgan o no sorteados. Los *nunca tomadores* deciden no participar en cualquiera de las contingencias. Los *desafiantes* hacen siempre lo contrario.

Mostraremos la estrategia estándar en la literatura para estudiar la relación entre el instrumento propuesto z y las variables d e y , con el fin de identificar el efecto tratamiento promedio *ATE*. Observando los valores de d condicionados a lo que obtengamos de z , se puede descomponer a d en dos variables *dummy*, d_1 y d_0 , las cuales son contingentes a los valores de z . Ambas variables toman el valor de 1 si el individuo participa en el programa y 0 si no participa. Aunque aparentemente ambas *dummies* sean iguales, en la realidad solo son observables parcialmente. Si participa en el programa, entonces vemos que $d_1 = 1$, si no participa en el programa observaremos que $d_0 = 0$.

Esta relación entre d y z se puede modelar para cualquier unidad i como:

$$d_1 = (1 - z_i)d_{0i} + z_i d_{1i}$$

A su vez, la variable resultado y se relaciona con d mediante la ecuación:

$$y_i = (1 - d_i)y_{0i} + d_i y_{1i} = y_{0i} + d_i(y_{1i} - y_{0i})$$

Donde claramente $y_i = y_{0i}$ si $d_i = d_{0i}$, e $y_i = y_{1i}$ si $d_i = d_{1i}$. Reemplazando la ecuación de d_i en y_i de la página anterior y con un poco de álgebra se obtiene (omitiendo el subíndice i):

$$y_1 = y_0 + d_0(y_1 - y_0) + z(d_1 - d_0)(y_1 - y_0)$$

En la literatura estadística se suele asumir que z es independiente de y_1 , y_0 , d_1 y d_0 , con lo cual:

$$E(y | z = 1) = E(y_0) + E(d_0(y_1 - y_0)) + E((d_1 - d_0)(y_1 - y_0))$$

$$E(y | z = 0) = E(y_0) + E(d_0(y_1 - y_0))$$

Luego, comparando la esperanza de los dos resultados dado que ocurre algún valor específico de z obtenemos:

$$E(y | z = 1) - E(y | z = 0) = E((d_1 - d_0)(y_1 - y_0)) = E(d_1 - d_0) E(y_1 - y_0 | d_1 - d_0)$$

Como $d_1 - d_0$ tiene tres posibles resultados: 1, 0 y -1:

$$E(y | z = 1) - E(y | z = 0) = 1 \cdot E(y_1 - y_0 | d_1 - d_0 = 1) \cdot \Pr(d_1 - d_0 = 1) + 0 \cdot E(y_1 - y_0 | d_1 - d_0 = 0) \cdot \Pr(d_1 - d_0 = 0) - 1 \cdot E(y_1 - y_0 | d_1 - d_0 = -1) \cdot \Pr(d_1 - d_0 = -1)$$

Si asumimos que $d_1 \geq d_0$, supuesto conocido como «monotonicidad», entonces $\Pr(d_1 - d_0 = -1) = 0$, con lo cual se elimina el tercer término de la última ecuación. Con ello la expresión se reduce a:

$$E(y | z = 1) - E(y | z = 0) = E(y_1 - y_0 | d_1 - d_0 = 1) \cdot \Pr(d_1 - d_0 = 1)$$

Dados los tres valores mencionados que puede tomar $d_1 - d_0$, la probabilidad se puede desmembrar de la siguiente forma:

$$\begin{aligned}\Pr(d_1 - d_0) &= E(d_1 - d_0) = E(d_1) - E(d_0) = E(d | z = 1) - E(d | z = 0) \\ &= \Pr(d = 1 | z = 1) - \Pr(d = 1 | z = 0)\end{aligned}$$

Aquí es necesario hacer el supuesto que $\Pr(d = 1 | z = 1) \neq \Pr(d = 1 | z = 0)$, el cual tiene sentido si el instrumento z afecta a d .³⁷ Reemplazando esta última ecuación en la penúltima y despejando tenemos:

$$E(y_1 - y_0 | d_1 - d_0 = 1) = \frac{E(y | z = 1) - E(y | z = 0)}{E(d | z = 1) - E(d | z = 0)} = \frac{\text{Efecto de } z \text{ sobre } y}{\text{Efecto de } z \text{ sobre } d}$$

Se puede comprobar³⁸ que:

$$\frac{E(y | z = 1) - E(y | z = 0)}{E(d | z = 1) - E(d | z = 0)} = \frac{\text{cov}(y, z)}{\text{cov}(d, z)} = \beta_d$$

En donde vemos que el rol de z ha sido de identificar el efecto de la variación de d sobre la variación de y (el parámetro β_d). Sin embargo, este valor solo mide la ganancia promedio de los *cumplidores* pues para ellos $d_1 - d_0 = 1$. No se puede calcular el efecto de los *siempre tomadores* pues no se observa variabilidad en su conducta ni el efecto de los *nunca tomadores*. Debido a que solo se está identificando el efecto en un subgrupo de la población, a este impacto se le llama efecto tratamiento promedio local (*LATE*)³⁹.

El estimador de variables instrumentales de β_d (tomando como instrumentos a z y la constante de unos) es el análogo muestral de la razón de covarianzas. Este es:

$$\hat{\beta}_d^{IV} = \frac{n \sum y_i z_i - \sum y_i \sum z_i}{n \sum d_i z_i - \sum d_i \sum z_i}$$

Otro estimador de variables instrumentales que es el análogo muestral de la razón de diferencias de esperanzas condicionales (también conocido como el estimador de Wald) que muestran Angrist, Imbens y Rubin (1996) es:

$$\hat{\beta}_d^{Wald} = \left(\frac{\sum y_i z_i}{\sum z_i} - \frac{\sum y_i (1 - z_i)}{\sum (1 - z_i)} \right) \cdot \left(\frac{\sum d_i z_i}{\sum z_i} - \frac{\sum d_i (1 - z_i)}{\sum (1 - z_i)} \right)^{-1} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{d}_1 - \bar{d}_0}$$

Donde el numerador de la expresión anterior es el estimador *intention-to-treat* del efecto de z sobre y , y el denominador el estimador *intention-to-treat* del efecto de z sobre d . Tanto $\hat{\beta}_d^{IV}$ como $\hat{\beta}_d^{Wald}$ son estimadores IV consistentes de *LATE*, la diferencia promedio de los *cumplidores*.

³⁷ Es la condición de relevancia que debe cumplir el instrumento.

³⁸ Lee (2005), 37.

³⁹ Por *Local Average Treatment Effect* en inglés

Heckman (1996) critica la validez del instrumento propuesto por Angrist, Imbens y Rubin, señalando que es plausible que exista alguna relación entre el resultado de la lotería y los resultados potenciales. Tal relación ocurre en el momento que aquellas personas que no fueron sorteadas pero que desean presentarse voluntariamente al servicio perciben una ganancia por hacerlo, basándose tal ganancia en información privada o características no observables. En tal sentido, el instrumento propuesto violaría la condición de exclusión del método.

Asimismo, Heckman (1996) muestra que el efecto tratamiento medio sobre los tratados *ATE* puede ser identificado usando variables instrumentales usando supuestos más débiles que los usados por Angrist, Imbens y Rubin (1996). Si llamamos $u_0 \equiv y_0 - E(y_0 | x)$ donde x son características observables de los individuos, entonces asumiendo la *independencia en medias condicionales* $E(u_0 | x, z) = 0$ y $E(y_1 - y_0 | x, z, d = 1) = E(y_1 - y_0 | x, d = 1)$, se puede comprobar que:

$$ATE | x = E(y_1 - y_0 | x, d = 1) = \frac{E(y | x, z = 1) - E(y | z = 0)}{\Pr(d = 1, x, z = 1) - \Pr(d = 1, x, z = 0)}$$

Donde solo se requiere que el denominador sea distinto de cero. Reemplazando estos términos poblacionales por análogos muestrales produce el estimador de variables instrumentales. Nótese que este resultado no requiere que z sea independiente de y_1 , y_0 , d_1 y d_0 , ni es necesario el supuesto de «monotonicidad».

Existen algunas aplicaciones interesantes del método de variables instrumentales en el análisis de inferencia causal. Por ejemplo, Angrist y Krueger (1991) analizan el efecto de los años de educación sobre los ingresos en Estados Unidos, en donde los años de educación son un regresor endógeno. El instrumento utilizado por estos autores es el trimestre de nacimiento, el cual es usado como predictor de la cantidad de años que un estudiante finalmente estudiará, debido a los reglamentos y leyes en el sistema educativo norteamericano.

En el estudio mencionado sobre la regla de las Maimonides, Angrist y Lavy (1999) encuentran que el tamaño de la clase no coincide exactamente con lo que predice la regla de los Maimonides debido a la presencia de cierta variabilidad en la determinación del número de estudiantes por aula. Sin embargo, utilizando a la regla de los Maimonides como un instrumento obtiene estimadores consistentes del efecto del tamaño de la clase sobre el rendimiento escolar en Israel.

En un ejemplo de la aplicación de esta técnica para tratamientos binarios, Schady y Araujo (2008) estudian el impacto del programa *Bono de Desarrollo Humano* de Ecuador —el cual es un programa de transferencias condicionadas de dinero— sobre la asistencia a la escuela. Si bien el programa se asignó en forma aleatoria entre la población objetivo debido a problemas presupuestarios, en la práctica hubo problemas de cumplimiento de lo establecido por el sorteo, existiendo autoselección en el tratamiento. Para superar este problema, los autores estiman el efecto del programa por variables instrumentales, considerando a la asignación aleatoria como un instrumento.

Pese a que el método de variables instrumentales es sumamente atractivo, se debe tener en cuenta que su eficacia se cumple solo si los instrumentos elegidos son válidos, es decir están correlacionados con el regresor endógeno (condición de relevancia) pero no lo están con el término de perturbación u (condición de exogeneidad o de exclusión⁴⁰). En caso que no se cumplan estas condiciones, los estimadores de variables instrumentales son sesgados e inconsistentes, y muestran sesgos bastante grandes. Por ejemplo, Bound, Jaeger y Baker (1995) critican los resultados de Angrist y Krueger (1991) al comprobar que la correlación entre el trimestre de nacimiento y los años de educación es muy débil, pudiendo entonces existir enormes sesgos pese al gran número de observaciones usado por Angrist y Krueger. Para una explicación más detallada del problema de instrumentos débiles, véase Staiger y Stock (1997), Stock y Yogo (2002) y Stock, Wright y Yogo (2002). Un resumen del problema y de las pruebas estadísticas que se pueden utilizar para la verificación de la validez de los instrumentos está en García (2008).

7. DISEÑO DE REGRESIÓN DISCONTINUA DIFUSA

Volviendo al trabajo de Angrist y Lavy sobre la regla de las Maimonides, los autores proponen que el impacto de la variación abrupta del tamaño de clase sobre el rendimiento escolar puede medirse localmente alrededor de la discontinuidad. Sin embargo, en algunas escuelas la matrícula en un determinado grado no determina exactamente si las aulas serán partidas o no, ni que las aulas divididas tengan el mismo número de alumnos, existiendo casos en donde no se llevó a cabo la partición o en donde la partición se realizó antes de llegar a lo estipulado por la regla. En ese sentido, el diseño de regresión discontinua aguda discutido en la sección 4 podría no funcionar.

Esta misma situación podría repetirse en otros programas en donde la asignación del tratamiento (recibirlo o no) dependa del valor de un indicador continuo. La heterogeneidad en las respuestas de los individuos ante la asignación del tratamiento puede generar problemas de no cumplimiento del mismo. Por ejemplo, en un programa de crédito educativo para alumnos que alcancen un rendimiento dado τ , habrá algunos alumnos con rendimiento superior al umbral que no solicitarán crédito así como habrá individuos por rendimiento ligeramente por debajo de τ , que podrían solicitar y recibir crédito si su situación familiar fuera muy crítica. También podría ocurrir que el comité evaluador de crédito podría ser flexible en algunos casos bajo criterios no controlados por el investigador.

En términos más generales y continuando con la discusión de la sección 4, suele ocurrir en algunos casos que la variable asignadora X (aquella cuyos valores determinarán la asignación del beneficio del programa, como la matrícula por grado en el ejemplo mencionado) no determina exactamente la participación o no en el programa aunque podría alterar la

⁴⁰ Es decir, no puede haber una relación directa entre el instrumento y la variable resultado. Solo puede haber una relación indirecta a través del regresor endógeno.

probabilidad de que participe en el mismo. En este caso diremos que la participación en el tratamiento es endógena, en el sentido que depende de la decisión de los agentes participantes.

En casos como el descrito ocurrirá que:

$$\Pr(d_i = 1 | X_i) = \begin{cases} \delta_1(X_i) & \text{si } X_i \geq c \\ \delta_0(X_i) & \text{si } X_i < c \end{cases}$$

Donde $\delta_1(X_i) \neq \delta_0(X_i)$ y $\delta_1(X_i) > \delta_0(X_i)$. Con esto decimos que para los individuos que están a la derecha de c es más probable que obtengan tratamiento que aquellos que están a la izquierda de c .

Tomando los gráficos que muestran Imbens y Lemieux (2007), en el gráfico 3 se muestra la discontinuidad en la probabilidad de recibir tratamiento y el efecto de esta diferencia sobre los resultados observables pueden apreciarse en el gráfico 4.

Gráfico 3
Discontinuidad en la probabilidad de recibir tratamiento

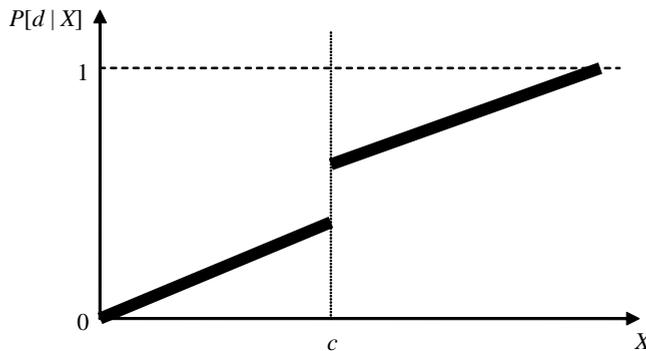
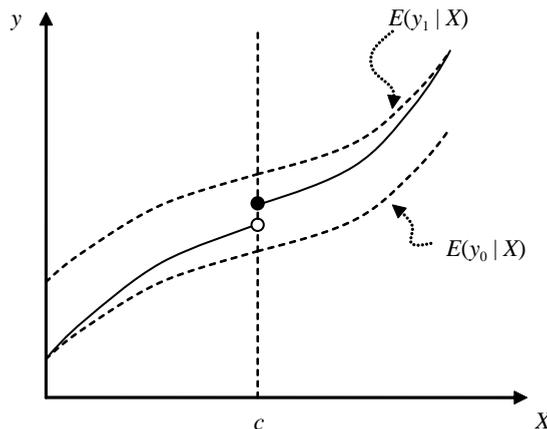


Gráfico 4
Ilustración de una regresión discontinua difusa



La consecuencia de esto en el modelo es que tendremos endogeneidad en el tratamiento d_i , al ser este no aleatorio sino dependiente de variables no observables:

$$y_i = \beta_d d_i + g(X_i) + u_i$$

En donde ahora d_i está correlacionado con u_i , pero no X_i .

Bajo el supuesto que $g(X_i)$ es continua y dado que $\lim_{X \rightarrow c} E(u | X) = 0$, entonces podemos tomar el límite por la derecha del esperado de la expresión anterior:

$$\lim_{X \downarrow c} E(y | X) = \beta_d \lim_{X \downarrow c} E(d | X) + \lim_{X \downarrow c} g(X) + \lim_{X \downarrow c} E(u | X)$$

Mientras que el límite por la izquierda es:

$$\lim_{X \uparrow c} E(y | X) = \beta_d \lim_{X \uparrow c} E(d | X) + \lim_{X \uparrow c} g(X) + \lim_{X \uparrow c} E(u | X)$$

Luego:

$$\lim_{X \downarrow c} E(y | X) - \lim_{X \uparrow c} E(y | X) = \beta_d \left(\lim_{X \downarrow c} E(d | X) - \lim_{X \uparrow c} E(d | X) \right)$$

Despejando:

$$\beta_d = \frac{\lim_{X \downarrow c} E(y | X) - \lim_{X \uparrow c} E(y | X)}{\lim_{X \downarrow c} E(d | X) - \lim_{X \uparrow c} E(d | X)}$$

El cual es el valor de *ATE*.

Siendo cuidadosos con lo que identifica este estimador, analicemos con cuidado lo que dice el denominador de la expresión anterior. Nótese que $E(d | X)$ indica el tratamiento esperado para cada valor de X . Supongamos que tenemos individuos heterogéneos en la población, en donde cada individuo tendrá una respuesta distinta de participación en el tratamiento ante un valor de su propio X_i . Los grupos mencionados en la sección 6 se definen en el contexto actual como:

Cumplidores: ($d_i = 1$ si $X_i \geq c$, $d_i = 0$ si $X_i < c$) Individuos que aceptan el tratamiento si su X_i se ubica a la derecha de c , no lo toman si X_i está a la izquierda de c .

Siempre tomadores: ($d_i = 1$ si $X_i \geq c$, $d_i = 1$ si $X_i < c$) Siempre participan.

Nunca tomadores: ($d_i = 0$ si $X_i \geq c$, $d_i = 0$ si $X_i < c$) Nunca participan.

En el caso de los *siempre tomadores* y los *nunca tomadores*, el denominador es cero pues no hay variabilidad en d para X alrededor de c . Por lo tanto, el *ATE* mencionado arriba, al igual que el estimador de variables instrumentales, solo está identificado para los individuos *cumplidores* alrededor del punto c .

En la aplicación práctica del método de regresión discontinua aguda o difusa, hay varias cuestiones a tomar en cuenta. En primer lugar se debería distinguir a qué caso de regresión discontinua corresponde el problema que estamos analizando. En esta etapa, un análisis gráfico de los datos suele ser de bastante utilidad. Algo muy importante que se debe verificar antes de empezar es que el salto en la variable resultado y se deba únicamente a los valores de X alrededor del umbral. Si X provoca saltos en otros determinantes de y , se distorsionaría el efecto del tratamiento. Algo que debe observarse también es que la distribución de las observaciones de X a ambos lados de c deberían ser simétricas. Si se observa una discontinuidad o fuertes asimetrías, se podría pensar que los individuos han manipulado sus valores de X con el fin de estar a un lado de c (para recibir o no el tratamiento). Si esto ocurre, se invalidaría el supuesto de intercambiabilidad de los individuos alrededor de c .

El método también presenta algunas limitaciones importantes. Una de ellas tiene que ver con la validez externa. Tanto el diseño de regresión discontinua aguda como difusa identifican al *ATE* únicamente alrededor de $X = c$. Incluso el diseño difuso se limita a una subpoblación aún más pequeña, los cumplidores. Cualquier extrapolación de estos resultados a otras subpoblaciones debe hacerse con cuidado. Otra limitación importante es el número de observaciones con que se cuenta alrededor del corte. Se requiere de muchas observaciones para tener estimadores confiables y precisos.

8. EL MÉTODO DE DIFERENCIAS EN DIFERENCIAS

Como se mencionó en la sección 5, el principal problema del método de pareo en la estimación del *ATE* es que no puede controlar las características no observables de los individuos, con lo cual existe un serio riesgo de sesgo en la estimación de este valor. Sin embargo, veremos en esta sección que podemos plantear un método que —bajo ciertos supuestos— es capaz de remover aquel componente no observable de los datos con el fin de tener estimaciones confiables. En síntesis, apoyándonos en la existencia de datos de panel de los individuos antes y después de recibir el tratamiento y asumiendo que las características no observables son invariantes en el tiempo podemos obtener estimaciones confiables del efecto tratamiento.

El tradicional método de diferencias en diferencias es un refinamiento del método de diferencias de Rubin de la sección 2.1 considerándose no solo la diferencia promedio de los resultados entre los individuos de los grupos B y C sino también la diferencia de la variable resultado antes y después del tratamiento. La idea de este procedimiento está en que se pretende eliminar cualquier componente sistemático y común a ambos grupos que vaya cambiando en el tiempo, el cual podría distorsionar el efecto del programa si se pretende medirlo como la diferencia de los resultados pos-tratamiento. Asimismo, la diferencia también puede eliminar cualquier otro componente individual no observable

de cada grupo. De este modo la diferencia en diferencia es una estrategia identificadora del efecto tratamiento promedio como mostramos a continuación.

Los resultados potenciales en este contexto dependerán no solo del individuo sino del tiempo, y_{jt} , donde $j = 0, 1$ muestra la exposición potencial o no al tratamiento —al igual que en la sección 2—, $t = 0, 1$ indica el tiempo, donde 0 es el periodo antes del tratamiento y 1 después del tratamiento, e i indica a la unidad i . Con el fin de evitar una notación engorrosa, hacemos $y_{jt} = y_t^j$ en donde se omitirá el subíndice i . Nótese que el indicador del resultado potencial se encuentra ahora como un superíndice. Por su parte, el resultado observado es $y_{it} = y_t$. Para la variable del tratamiento se escribirá $d_{it} = d_t$. Esta variable no solo indica la recepción del tratamiento en cada periodo para la unidad i sino que también indica si estamos hablando del grupo beneficiario ($d_1 = 1$) o del grupo no beneficiario ($d_1 = 0$) en cualquiera de los dos periodos.

El efecto tratamiento promedio sobre los tratados en el periodo 1 se define como:

$$\delta_T = ATET = E(y_1^1 | d_1 = 1) - E(y_1^0 | d_1 = 1)$$

Donde puede verse que el segundo término $E(y_1^0 | d_1 = 1)$ no es observable pues es el resultado promedio que hubieran obtenido los beneficiarios en caso no hubieran recibido el tratamiento.

Se podría pensar en el grupo no beneficiario como el contrafactual del beneficiario. Sin embargo, en este contexto la diferencia de medias de los beneficiarios y no beneficiarios en el periodo 1 (pos-tratamiento) no identifica al *ATET*. Veamos:

$$E(y_1 | d_1 = 1) - E(y_1 | d_1 = 0) = E(y_1^1 | d_1 = 1) - E(y_1^0 | d_1 = 0)$$

Sumando y restando $E(y_1^0 | d_1 = 1)$ se obtiene:

$$E(y_1 | d_1 = 1) - E(y_1 | d_1 = 0) = \delta_T + E(y_1^0 | d_1 = 1) - E(y_1^0 | d_1 = 0)$$

La expresión $E(y_1^0 | d_1 = 1) - E(y_1^0 | d_1 = 0)$ muestra la diferencia en los resultados potenciales en ausencia de tratamiento entre los dos grupos en el periodo 1. Si se mantiene el supuesto que el tratamiento es independiente en medias condicionales con y^0 , entonces tal diferencia sería igual a cero. Si no se cumple, entonces la diferencia en medias post-tratamiento no identifica el *ATE*. Esa diferencia captura aquel componente individual que no está balanceado entre los dos grupos.

Análogamente, la diferencia en medias en el periodo cero para los dos grupos es:

$$E(y_0 | d_1 = 1) - E(y_0 | d_1 = 0) = E(y_0^0 | d_1 = 1) - E(y_0^0 | d_1 = 0)$$

La cual debería ser cero ante aleatorización del tratamiento o menos rigurosamente cuando $y_0 \perp d$. Cuando no se cumple esto, captura las diferencias en el resultado potencial 0 para ambos grupos en el periodo 0.

Si tales diferencias entre los beneficiarios y no beneficiarios se mantienen en $t = 0$ y $t = 1$, entonces ocurrirá que:

$$E(y_1^0 | d_1 = 1) - E(y_1^0 | d_1 = 0) = E(y_0^0 | d_1 = 1) - E(y_0^0 | d_1 = 0) \equiv \Delta d$$

Luego la diferencia de la diferencia identifica al *ATET*.

$$[E(y_1 | d_1 = 1) - E(y_1 | d_1 = 0)] - [E(y_0 | d_1 = 1) - E(y_0 | d_1 = 0)] = \delta_T$$

Otra forma de obtener el mismo resultado es la siguiente. La diferencia en resultados observables *antes y después* del tratamiento para el grupo $B (d_1 = 1)$ es:

$$E(y_1 | d_1 = 1) - E(y_0 | d_1 = 1) = E(y_1^1 | d_1 = 1) - E(y_0^1 | d_1 = 1)$$

Sumando y restando $E(y_1^0 | d_1 = 1)$ se obtiene:

$$\begin{aligned} E(y_1 | d_1 = 1) - E(y_0 | d_1 = 1) &= \delta_T + E(y_1^0 | d_1 = 1) - E(y_0^1 | d_1 = 1) \\ &= \delta_T + E(y_1^0 | d_1 = 1) - E(y_0^0 | d_1 = 1) \end{aligned}$$

En donde se ha tomado en cuenta que antes del tratamiento, $E(y_0^0 | d) = E(y_0^1 | d)$, la expresión $E(y_1^0 | d_1 = 1) - E(y_0^0 | d_1 = 1)$ captura el efecto temporal sobre y^0 para los beneficiarios. No hay ninguna razón para asumir que tal efecto es cero, ni siquiera en experimentos aleatorios.

Para el grupo $N (d_1 = 0)$, la diferencia de medias *antes y después* es:

$$E(y_1 | d_1 = 0) - E(y_1 | d_1 = 0) = E(y_1^0 | d_1 = 0) - E(y_0^0 | d_1 = 0)$$

Donde esta diferencia muestra el efecto del tiempo sobre y^0 para los no beneficiarios. Nada garantiza que este efecto temporal sobre y^0 sea igual a aquel de los beneficiarios. En este punto se requiere asumir que tales efectos temporales son iguales para B y N (ambos siguen la misma tendencia):

$$E(y_1^0 | d_1 = 1) - E(y_0^0 | d_1 = 1) = E(y_1^0 | d_1 = 0) - E(y_0^0 | d_1 = 0) \equiv \Delta t$$

Luego la diferencia de la diferencia identifica el *ATET*:

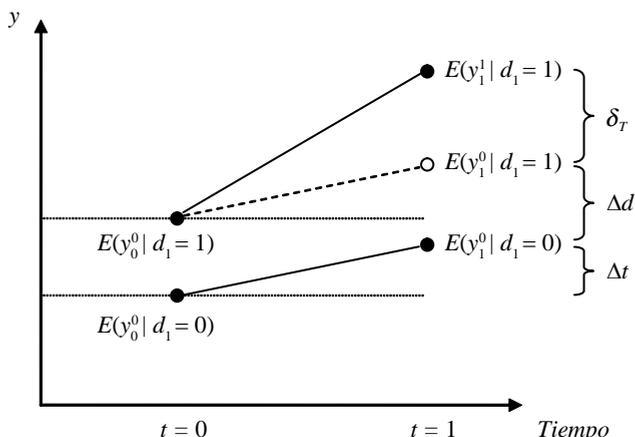
$$[E(y_1 | d_1 = 1) - E(y_0 | d_1 = 1)] - [E(y_1 | d_1 = 0) - E(y_0 | d_1 = 0)] = \delta_T$$

Utilizando análogos muestrales, un estimador consistente de *ATET* es:

$$\hat{\delta}_T = (\bar{y}_{i \in B, t=1} - \bar{y}_{i \in B, t=0}) - (\bar{y}_{i \in N, t=1} - \bar{y}_{i \in N, t=0})$$

En el gráfico 5 se puede apreciar el efecto de la doble diferenciación. En la figura se toma como base al periodo $t = 0$ en donde se cuenta con observaciones de y para ambos grupos, y se tiene información en el periodo $t = 1$ después de la aplicación del tratamiento, también para ambos grupos. Los puntos negros indican valores realizados mientras que el punto blanco indica el resultado potencial no observable.

Gráfico 5
Ilustración del método de diferencias en diferencias



Puede notarse que el supuesto acerca del mismo efecto temporal en y^0 se cumple al ser paralelas las rectas $[E(y_0^0 | d_1 = 0), E(y_1^0 | d_1 = 0)]$ y la recta $[E(y_0^0 | d_1 = 1), E(y_1^0 | d_1 = 1)]$. Si tal paralelismo no se cumple, el estimador de diferencias en diferencias sería un estimador incorrecto de *ATE*.

El gráfico anterior muestra claramente que el estimador de diferencias en diferencias es el *ATE* y no el *ATE*. Es decir el cálculo no es válido externamente para otros grupos distintos a los beneficiarios.

En términos de regresiones, se puede obtener el estimador de diferencias en diferencias de la regresión:

$$y_{it} = \beta_d d_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

Donde y_{it} es el valor observado del resultado, μ_i muestra un efecto individual (efecto fijo) no observable que afecta a la variable resultado, λ_t es un componente temporal no observable que genera el «efecto periodo» y ε_{it} es un término de error de media cero, varianza condicional constante y no correlacionado con ninguna de las demás variables ni con ningún otro error. Se puede comprobar sin mayor dificultad que la diferencia en diferencia:

$$E(y_{i1} - y_{i0} | d_1 = 1) - E(y_{i1} - y_{i0} | d_0 = 0)$$

Es igual al parámetro β_d , el cual es el efecto causal que se desea estimar.

No hemos mencionado nada acerca del control en variables x . Sin embargo, se puede extender el análisis condicionando a diferentes valores de estas variables de control. Bajo la existencia en un soporte común en las características x , y bajo el supuesto de iguales tendencias de y_{it}^0 en los tratados y no tratados condicional a valores de x , se puede emplear el método de pareo para calcular el *ATE*. Una metodología usada en datos

observacionales en donde no se aprecie una clara aleatorización del tratamiento es el *método de pareo en diferencias en diferencias condicional* (Heckman et al. 1998). Este estimador es (sin omitir el subíndice i del individuo o unidad de análisis):

$$AT\hat{E}T = \frac{1}{n_B} \sum_{i \in B \subset CS} \left((y_{i1}^1 - y_{i0}^0) - \sum_{i \in B \subset CS} \omega(i, k) (y_{k1}^1 - y_{k0}^0) \right)$$

Donde A_i es el grupo comparable a la unidad i que tiene las características comunes con los beneficiarios, y los ponderadores $\omega(i, k)$ son los que se presentaron en la sección 5.1.

Existen numerosas aplicaciones del método de diferencias en diferencias en estudios observacionales y también en experimentales⁴¹. Por ejemplo, Chong y Galdo (2006) estudian el impacto del programa de capacitación juvenil Projooven sobre los salarios ganados y utilizan el estimador de pareo en diferencias en diferencias condicional, aunque los autores ajustan el estimador con el fin de analizar variaciones en la calidad del programa.

En estudios que trabajan con modelos de regresión lineal con efectos fijos, Di Tella y Schargrodsky (2004) calculan el impacto de la presencia policial en la reducción del crimen, tomando datos de un experimento natural surgido de un atentado terrorista en la ciudad de Buenos Aires. Tras el atentado, se reforzó la seguridad alrededor de los locales de 45 instituciones judías y musulmanas en la ciudad. Aplicando el método de diferencias en diferencias a través de regresiones con datos de panel, encuentran un efecto significativo de una mayor presencia policial en el número promedio de autos robados en comparación con vecindarios de similares características socioeconómicas pero que no contaban con presencia policial adicional. En otro documento, Galiani, Gertler y Schargrodsky (2005) estudian el impacto de la privatización en el suministro de agua en Argentina sobre la mortalidad infantil. Habiendo verificado que la mortalidad infantil mantenía una misma tendencia antes de la aplicación de la política (lo cual sustentaría el supuesto de iguales tendencias de los resultados potenciales sin tratamiento en los grupos beneficiarios y no beneficiarios), la regresión con efectos fijos encuentra que la mortalidad infantil se redujo como consecuencia de la privatización.

9. CONSIDERACIONES FINALES

El objetivo de este documento ha sido mostrar los métodos más populares utilizados para identificar y estimar el efecto causal de una política con tratamiento binario, sobre un resultado. El énfasis se ha puesto en términos de la identificación de dicho efecto, la cual dependerá de una serie de supuestos que se ajustan a los datos que se tiene disponible.

⁴¹ Se puede comprobar que en el caso de estudios experimentales existe una ganancia en eficiencia cuando se utiliza el estimador de diferencias en diferencias, en comparación con la diferencia simple de los beneficiarios y el grupo de control en el periodo post-tratamiento.

A lo largo del documento, los diferentes temas han sido presentados desde la óptica del economista, teniendo en cuenta que el objetivo final es la estimación de efectos causales. Para tal fin, el conocimiento sobre la conducta de los agentes y los modelos económicos que han originado los datos son imprescindibles, pues ellos configuran los problemas de datos (tales como la selección, y la endogeneidad del tratamiento) que los métodos buscan resolver en diferentes contextos. De ninguna manera este documento pretende difundir las técnicas de evaluación de impacto mediante un «empirismo ciego» carente de una teoría de la conducta de los agentes económicos.

La elección final del método debe basarse en la problemática de estudio, pero debe en todo momento tenerse en cuenta los supuestos bajo los cuales tal método realmente identifica al efecto tratamiento promedio buscado.

REFERENCIAS

ANGRIST, Joshua D.

1990 «The effect of veteran benefits on veterans' education and earnings». National Bureau of Economic Research, Working Paper N° 3492.

ANGRIST, Joshua, Eric BETTINGER, Erik BLOOM, Elizabeth KING y Michael KREMER

2002 «Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment». *The American Economic Review*, Vol. 92, No. 5, 1535-1558.

ANGRIST, Joshua, Guido W. IMBENS y Donald B. RUBIN

1996 «Identification of causal effects using instrumental variables». *Journal of the American Statistical Association*, Vol. 91, N° 434, 444-455.

ANGRIST, Joshua D. y Alan B. KRUEGER

1991 «Does compulsory school attendance affect schooling and earnings?». *Quarterly Journal of Economics*, Vol. 106, 979-1014.

2001 «Instrumental variables and the search for identification: from supply and demand to natural experiments». *The Journal of Economic Perspectives*, Vol. 15, N° 4, 69-85.

ANGRIST, Joshua D. y Victor LAVY

1999 «Using Maimonides' rule to estimate the effect of class size on scholastic achievement». *The Quarterly Journal of Economics*, Vol. 114 N°2, 533-575.

2002 «The effect of high school matriculation awards: evidence from randomized trials». National Bureau of Economic Research Working Paper N° 9389. Disponible en: <http://www.nber.org/papers/w9389>.

BANERJEE, Abhijit, Shawn COLE, Esther DUFLO y Leigh LINDEN

2004 «Remedying Education: Evidence from two randomized experiments in India». Mimeo, MIT.

BARRERA-OSORIO, Felipe, Leigh L. LINDEN y Miguel URQUIOLA

2007 «The effects of user fee reductions on enrollment: evidence from a quasi-experiment». Mimeo, Columbia University. Disponible en: <http://www.columbia.edu/~ll2240/Gratuidad%20Draft%202007-01.pdf>

BOUND, John, David A. JAEGER y Regina M. BAKER

1995 «Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak.» *Journal of the American Statistical Association*, Vol. 90, N° 430, 443-450.

BURGA, Cybele

2003 «Re-evaluando PROJOVEN: propensity score matching y una evaluación paramétrica». Consorcio de Investigación Económica y Social (CIES). Disponible en: www.cies.org.pe/files/active/0/Pb0112.pdf.

CALIENDO, Marco y Sabine KOPEINIG

2005 «Some Practical Guidance for the implementation of propensity score matching». IZA Discussion Paper N° 1588. The Institute for the Study of Labor.

CAMPBELL, Donald T.

1969 «Reforms as experiments.» *American Psychologist*, N° 24, 409-429.

COCHRAN, William G.

1965 «The planning of observational studies of human populations». *Journal of the Royal Statistical Society. Series A (general)*, Vol.128, N°2, 234-255.

CHONG, Alberto y José GALDO

2006 «Training quality and earnings: the effects of competition on the provision of public-sponsored training programs.» Banco Interamericano de Desarrollo. Working Paper N° 555.

DEHEJIA, Rajeev H. y Sadek WAHBA

1999 «Causal effects in nonexperimental studies: reevaluating the evaluation of training programs». *Journal of the American Statistical Association*, Vol. 94, N° 448, 1053-1062.

2002 «Propensity Score-Matching Methods for Nonexperimental Causal Studies». *The Review of Economics and Statistics*, Vol. 84, N° 1, 151-161

DI NARDO, John y David S. LEE

2002 «The impact of unionization on establishment closure: A regression discontinuity analysis of representation elections». National Bureau of Economic Research, Working Paper N° 8993, junio. Disponible en: <http://www.nber.org/papers/w8993>.

DI TELLA, Rafael y Ernesto SCHARGRODSKY

2004 «Do police reduce crime? Estimates using the allocation of policy forces after a terrorist attack». *The American Economic Review*, Vol. 94. N°1, 115-133.

DONNER, Allan y Neil KLAR

2004 «Pitfalls of and controversies in cluster randomization trials». *American Journal of Public Health*, 94: 3; 416-422.

DONNER, Allan, Stephen BROWN y Penny BRASHER

1990 «A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989». *International Journal of Epidemiology*. Vol 19, N°4, 795-800.

GALIANI, Sebastian, Paul GERTLER y Ernesto SCHARGRODSKY

2005 «Water for life: the impact of the privatization of water services on child mortality.» *The Journal of Political Economy*, Vol. 113, N° 1, 83-120.

GARCÍA, Luis

2008 «IV estimation with weak instruments: an application to the determinants of school attainment in Peru». *Economía*, Vol. XXXI, N° 62, Lima.

GERTLER, Paul

2004 «Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment». *American Economic Review: Papers and Proceedings*, 94: 2; 336-334.

GOLDBERGER, A.

1972 «Structural Equations Methods in the Social Sciences». *Econometrica*, Vol. 40, N° 6, 979-1002.

HAAVELMO, Trygve

1943 «The statistical implications of a system of simultaneous equations». *Econometrica*, Vol. 11, N° 1, 1-12.

1944 «The probability approach in econometrics». *Econometrica*, Vol. 12, Supplement (Julio), iii-115.

HAHN, Jinyong, Petra E. TODD y Wilbert VAN DER KLAUW

2001 «Identification and Estimation of treatment effects with a regression-discontinuity design». *Econometrica*, Vol. 69, N°1, 201-209.

HECKMAN, James

1996 «Identification of causal effects using instrumental variables: comment». *Journal of the American Statistical Association*, Vol. 91, N° 434, 459-462.

2000 «Causal parameters and policy analysis in economics: a twentieth century retrospective». *The Quarterly Journal of Economics*, Vol. 115, 45-97.

2005 «The scientific model of causality». *Sociological Methodology*. Vol. 35, N°1, 1-97.

HECKMAN, James J., Hidehiko ICHIMURA y Petra E. TODD

1997 «Matching as an econometric evaluation estimator: evidence from evaluating a job training programme.» *The Review of Economic Studies*, Vol. 64, N° 4, Special Issue: Evaluation of Training and Other Social Programmes, 605-654.

1998 «Matching as an econometric evaluation estimator». *The Review of Economic Studies*, Vol. 65, N° 2 (Abril), 261-294.

HECKMAN, James, Hidehiko ICHIMURA, Jeffrey SMITH y Petra E. TODD

1998 «Characterizing selection bias using experimental data». *Econometrica*, Vol. 66, N° 5 (Septiembre), 1017-1098.

HODDINOTT, John y Emmanuel SKOUFIAS

2003 «The impact of ProgresA on food consumption». FCNF Discussion Paper N° 150. Washington: IFPRI.

HOLLAND, Paul W.

1986 «Statistics and causal inference». *Journal of the American Statistical Association*. Vol. 81, N° 396, Theory and Methods, 945-960.

IMBENS, Guido

2000 «The role of the propensity score in estimating dose-response functions». *Biometrika*, Vol. 87, N° 3 (Sep., 2000), 706-710.

IMBENS, Guido y Joshua D. ANGRIST

1994 «Identification and estimation of local average treatment effects». *Econometrica*, Vol. 62, N° 2 (Marzo 1994), 467-475.

IMBENS, Guido y Thomas LEMIEUX

2007 «Regression discontinuity designs: a guide to practice». National Bureau of Economic Research, Working Paper N° 13039. Disponible en: <http://www.nber.org/papers/w13039>.

JOFFE, Marshall M. y Paul R. ROSENBAUM

1999 «Propensity scores». *American Journal of Epidemiology*. Vol. 150, N°4, 327-333.

LALONDE, Robert J.

1986 «Evaluating the econometric evaluations of training programs with experimental data». *The American Economic Review*, Vol. 76, N° 4 (Sep.), 604-620.

LECHNER, Michael

2000 «An evaluation of public-sector-sponsored continuous vocational training programs in East Germany». *The Journal of Human Resources*, Vol. 35, N° 2 (Spring), 347-375.

2001 «Identification and estimation of causal effects of multiple treatment under the conditional independence assumption». Michael Lechner y Friedhelm Pfeiffer (Eds.). *Econometric Evaluation of Labour Markets Policies*. Heidelberg: Physica-Verlag.

2002 «Program heterogeneity and propensity score matching: an application to the evaluation of active labor market policies». *The Review of Economics and Statistics*, Vol. 84, N° 2 (May), 205-220.

LEE, Myoung-Jae

2005 *Micro-econometrics for policy, program, and treatment effects*. Oxford: Oxford University Press.

LEE, David S., Enrico MORETTI y Matthew J. BUTLER

2004 «Do Voters Affect or Elect Policies? Evidence from the U. S. House». *The Quarterly Journal of Economics*, Vol. 119, N° 3 (Aug.), 807-859.

MADDALA, G.S.

1982 *Limited, dependent and qualitative variables in Econometrics*. Cambridge: Cambridge University Press.

NEYMAN, Jerzy Splawa

1990 «On the application of probability theory to agricultural experiments. Essay on principles. Section 9». *Statistical Science*, Vol. 5, N° 4, 465-480. Traducción del original publicado en *Roczniki Nauk Rolniczych* Tomo X (1923), 1-51 (*Annals of Agricultural Sciences*).

PEARL, Judea

2000 *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

2009 «Causal inference in statistics: an overview». *Statistics Surveys*, Vol. 3, 96-146.

QUANDT, R. E.

1958 «The estimation of the parameters of a linear regression system obeying two separate regimes». *Journal of the American Statistical Association* 53 (284); 873-880.

ROSENBAUM, P.R.

2009 *Design of observational studies*. New York: Springer.

ROSENBAUM, P.R. y Donald B. RUBIN

1983 «The central role of the propensity score in observational studies for causal effects». *Biometrika*, Vol. 70, N°1, 41-55.

ROY, D.

1951 «Some thoughts on the distribution of earnings». *Oxford Economic Papers*, New Series, Vol. 3, N° 2 (Jun.), 135-146.

RUBIN, Donald B.

1974 «Estimating causal effects of treatments in randomized and nonrandomized experiments». *Journal of Educational Psychology* N° 66, 688-701.

SCHADY, Norbert y Maria Caridad ARAUJO

2008 «Cash transfers, conditions, and school enrollment in Ecuador». *Economía (Journal of the Latin American and Caribbean Economic Association)*. Vol. 8, N° 2 Spring.

STAIGER, Douglas y James H. STOCK

1997 «Instrumental variables regression with weak instruments». *Econometrica* 65, N° 3 (May), 557-586.

STOCK, James y Motohiro YOGO

2002 «Testing for weak instruments in linear IV regressions». NBER Technical Working Paper N° 284. Cambridge, MA.

STOCK, J., J. H. WRIGHT y M. YOGO

2002 «A survey of weak instruments and weak identification in generalized method of moments». *Journal of Business & Economic Statistics* 20, N°4 (October), 518-529.

STOCK, James y Mark WATSON

2003 *Introduction to Econometrics*. Boston: Addison Wesley.

THISTLETHWAITE, Donald L. y Donald T. CAMPBELL

1960 «Regression-Discontinuity Analysis: An alternative to the ex post facto experiment». *Journal of Educational Psychology* 51: 6; 309-317.

VINHA, Katja

2006 «A primer on propensity score matching estimators». Documento CEDE N° 2006-13. Universidad de los Andes.

WOOLDRIDGE, Jeffrey

2001 *Econometric Analysis of Cross Section and Panel Data*. London: The MIT Press.