tes en mejora de la formación sino también para los participantes de otros programas si hubieran participado en la mejora de la formación.

Por último, el Capítulo XI utiliza una muestra que enlaza datos de una encuesta de empleo con datos administrativos de deuda de los hogares para estimar el impacto causal de acceso al mercado de crédito sobre la formación de nuevos hogares entre los jóvenes. Para identificar cambios exógenos en el acceso al crédito utiliza una reforma en un programa portugués denominado “*Crédito Bonificado*” y que proporciona subsidios directos al endeudamiento de jóvenes con rentas medias y medias-bajas. Se encuentran dos resultados principales. El acceso al mercado de crédito aumenta la probabilidad de formar un hogar entre 31 y 54 puntos porcentuales. Asimismo, encuentran que las diferencias en el acceso al crédito pueden explicar en torno al 20% de la variación en tasas de salida del hogar. Los resultados también sugieren que los jóvenes se aseguran contra variaciones en el coste del crédito extendiendo la estancia en el hogar paterno.

II INFERENCIA CAUSAL Y EVALUACIÓN DE POLÍTICAS PÚBLICAS: METODOLOGÍA Y APLICACIONES

Cesar Alonso-Borrego
Raquel Carrasco
Universidad Carlos III (Madrid)

II.1 Introducción

El sector público suele dedicar gran cantidad de recursos a financiar programas sociales, como por ejemplo programas de formación para trabajadores desempleados o desfavorecidos, o mejoras educativas. En general, dichos programas tienen como finalidad influir positivamente sobre las perspectivas económicas y sociales de determinados colectivos. El objetivo de la evaluación de políticas consiste en estudiar la efectividad de dichas intervenciones públicas y diseñar políticas más eficaces y rentables desde el punto de vista social. Para ello, es crucial poder dar respuesta a la siguiente pregunta: ¿cuál es el efecto causal de una determinada política pública (o “tratamiento”) sobre la variable de interés?

Para dar respuesta a dicha pregunta, cabría pensar que bastaría comparar individuos sometidos al tratamiento con aquellos no sometidos a él. Sin embargo, dicha comparación está generalmente sesgada por problemas de selección endógena. Es decir, es posible que los individuos tratados y no tratados (por ejemplo, los que realizan cursos de formación en el empleo y los que no) difieran en los valores de las variables de interés (por ejemplo, el salario) por razones que ajenas y previas al tratamiento (es decir, al hecho de haber realizado un curso de formación o de

no haberlo realizado). Los métodos de inferencia causal estudian cómo estimar efectos causales en estas situaciones con objeto de evaluar los efectos de una determinada política pública. En las siguientes secciones se analizará la metodología econométrica de inferencia causal, así como la evidencia empírica más relevante.

II.1.1 Evaluación de políticas públicas: El problema de identificación

En la literatura de evaluación de políticas, se entiende por tratamiento la participación en alguno de los programas asociados a dichas políticas, mientras que por efecto causal se entiende la variación de la variable de interés que se debe exclusivamente al tratamiento, y no a otros factores. El concepto de causalidad es clave en este contexto. Para definirlo apropiadamente, debemos introducir algo de notación. Sea D_i una variable que indica si el individuo i ha recibido el tratamiento (por ejemplo, participación en un programa de formación), en cuyo caso tomará el valor 1, mientras que tomará el valor en caso contrario. Al colectivo de individuos que no han recibido el tratamiento se les denomina grupo de control. Sea Y_i el valor de la variable de interés para el individuo i (por ejemplo, el salario). Para definir causalidad, utilizaremos el concepto de resultados potenciales. Así, sea Y_{0i} el resultado potencial en ausencia del tratamiento para el individuo i , es decir, el resultado que el individuo i alcanzaría si no se viera sometido al tratamiento (por ejemplo, su salario potencial sin realizar el curso de formación). Asimismo, sea Y_{1i} el resultado potencial en presencia del tratamiento para el individuo i , es decir, el resultado que el individuo i alcanzaría si se viera sometido al tratamiento (por ejemplo, su salario potencial si realizara el curso de formación).

Dichos resultados potenciales se refieren a situaciones contra factuales. Para un individuo que de hecho no haya participado en el programa de formación, Y_{1i} representa el salario hipotético que hubiera obtenido si hubiera participado. Igualmente, para un participante en el programa, representa el salario que hubiera obtenido en caso de no haber participado en el programa. Sin embargo, para cada individuo no observamos ambos resultados potenciales (Y_{1i} , Y_{0i}), sino que únicamente observamos el resultado realizado, Y_i que podemos expresar como:

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

Una vez definidos los resultados potenciales, el efecto causal del tratamiento para el individuo i se define de forma natural como la diferencia entre ambos resultados potenciales,

$$Y_{1i} - Y_{0i}.$$

De esta forma, queda patente el problema fundamental de identificación (conocido como el problema de evaluación): dado que para el mismo individuo no podemos observar Y_{1i} e Y_{0i} no es posible calcular, para cada individuo, el efecto causal definido previamente.

El anterior problema se podría solucionar si hubiera homogeneidad entre todos los individuos. En ese caso, si por ejemplo (Y_{1i} , Y_{0i}) fuera constante entre individuos, se podría calcular la diferencia ($Y_{1i} - Y_{0i}$) para cada individuo mediante la comparación de los resultados objeto de estudio entre un individuo tratado y otro no tratado. Sin embargo, en general, habrá un alto grado de heterogeneidad en las respuestas y de características los individuos que participan y que no participan en programas públicos, por lo que no está garantizado que la anterior diferencia refleje exclusivamente el efecto de la política.

A pesar de que no es posible calcular efectos individuales del tratamiento, bajo ciertas condiciones se pueden estimar efectos medios. Por ejemplo, es posible estimar el llamado efecto medio del tratamiento (*Average Treatment Effect*, ATE):

$$\alpha_{ATE} = E[Y_1 - Y_0],$$

donde E representa el operador esperanza matemática.

También es posible estimar otro tipo de efectos medios, como por ejemplo el efecto medio sobre los individuos tratados, también llamado efecto medio del tratamiento seleccionado (SATE):

$$\alpha_{SATE} = E[Y_1 - Y_0 \mid D=1].$$

En cualquier caso, la comparación de los resultados de interés, , para individuos tratados y no tratados, no proporciona en general la respuesta que buscamos como evaluadores de políticas públicas, esto es, el efecto causal del tratamiento. La razón es que existe un sesgo de selección, es decir, los individuos no se distribuyen aleatoriamente entre tratados y no tratados, sino que suelen existir características inobservables para el analista que influyen en el hecho de recibir o no tratamiento, y que también tienen un efecto directo sobre el resultado de interés. En términos de la comparación de resultados medios para individuos tratados y no tratados, podemos ver fácilmente que:

$$\begin{aligned} E[Y \mid D=1] - E[Y \mid D=0] &= E[Y_1 \mid D=1] - E[Y_0 \mid D=0] \\ &= E[Y_1 - Y_0 \mid D=1] + \{E[Y_0 \mid D=1] - E[Y_0 \mid D=0]\}. \end{aligned}$$

El problema es que, en general, el término que hay entre llaves no será igual a cero, debido a problemas de selección. Por ejemplo, en el caso de participación en programas de formación, la mayoría de las veces los participantes en los mismos suelen ser individuos que están en una situación laboral más desfavorable, por lo que

podemos esperar que, incluso en ausencia de la participación en el programa, los salarios hipotéticos de los participantes fueran menores que los de los no participantes. Por ello, la simple comparación entre las ganancias de los individuos tratados y no tratados no aportará información sobre el efecto del tratamiento.

La *Inferencia Causal* es una rama de la Estadística que estudia la obtención de efectos causales, tanto a partir de datos experimentales como no experimentales. Para ello, básicamente, se utilizará la información sobre el mecanismo de asignación de los individuos al tratamiento.

II.1.2 Experimentos aleatorizados

En la literatura de estimación de efectos causales de un tratamiento, se considera que la situación ideal para un investigador es contar con datos procedentes de un experimento aleatorizado. Se trata de experimentos controlados en los que la asignación o no al tratamiento se decide aleatoriamente por el investigador. Esta situación, por construcción, asegura que los resultados potenciales no se ven contaminados por factores distintos al tratamiento. De hecho, en un experimento aleatorio ideal no es necesario controlar por el efecto de otras variables observables, a no ser que estemos interesados en obtener efectos diferenciados para grupos específicos de individuos.

No obstante, con este tipo de datos es importante controlar la validez de las estimaciones obtenidas. Dicha validez se clasifica en dos grupos:

— Validez Interna: Significa que es posible estimar el efecto del tratamiento para nuestra muestra particular. La validez interna no se cumplirá cuando existan diferencias entre individuos tratados y no tratados que afecten al resultado de interés y que no seamos capaces de controlar.

— Validez externa: Significa que es posible extrapolar nuestras estimaciones a otras poblaciones. La validez externa no se cumplirá cuando fuera del marco concreto de la evaluación el tratamiento tenga un efecto diferente.

En los experimentos aleatorios, las amenazas más comunes a la validez interna del mismo son la falta de aleatoriedad, la falta de cumplimiento con las normas o el protocolo del experimento, y el desgaste muestral no aleatorio (*attrition*), que se produce cuando un grupo de individuos deja de participar en el experimento por razones que están relacionadas con el resultado de interés (y por tanto, el desgaste es no aleatorio). Supongamos un experimento que trata de evaluar el efecto de programas de formación en el empleo sobre los salarios. Si un grupo de individuos abandonase aleatoriamente la muestra, este desgaste muestral no generaría problemas de sesgo. Sin embargo, si por ejemplo aquellos individuos de mayor sala-

rio abandonan voluntariamente la muestra, este desgaste muestral no aleatorio sí tendría efectos sobre el resultado de evaluación de la política en cuestión. En lo que respecta a la validez externa del experimento, las posibles fuentes de incumplimiento provienen de la no representatividad de la muestra y de la no representatividad del programa.

II.1.3 Estudios observacionales

En muchas ocasiones, no se dispone de datos procedentes de experimentos aleatorizados, lo que puede estar motivado por el elevado coste de acometer dicho experimento o por la prohibición legal de tratar de manera discriminatoria (asignando o no al tratamiento) a individuos que son iguales en lo que respecta a sus características relevantes. En todo caso, en ausencia de datos experimentales, los investigadores deben basar sus inferencias en datos observacionales, dando lugar a los llamados estudios observacionales. En este tipo de estudios, esperamos que los individuos tratados y no tratados difieran en características que afectan a la variable de interés, además de en su exposición o no al tratamiento. Si dichas características son observables, es posible tenerlas en cuenta de forma relativamente sencilla. Esta situación se conoce como “selección basada en observables” y da lugar a estimadores basados en el matching o emparejamiento y a estimadores basados en métodos de regresión.

No obstante, en muchas ocasiones tenemos buenas razones para pensar que los individuos tratados y no tratados son diferentes también en características inobservables para el analista. Dicha situación también puede solventarse si suponemos, por ejemplo, que observamos a los individuos tratados antes y después del tratamiento. En estas circunstancias, podríamos utilizar la información previa al tratamiento de los individuos tratados, construyendo el grupo de comparación a partir de dicha información previa y comparando a los individuos con ellos mismos después del tratamiento. El problema es que dicha comparación puede estar de nuevo contaminada por el efecto de determinados acontecimientos que se produzcan entre los dos periodos, antes y después de que se lleve a cabo el tratamiento. No obstante, se podría utilizar el grupo de individuos no sometidos al tratamiento para identificar la variación temporal en el resultado de interés que no es debida al tratamiento. El estimador de diferencias-en-diferencias se basa en esta idea.

En otras circunstancias, es posible identificar el efecto causal del tratamiento utilizando alguna variable determinada exógenamente (llamada normalmente instrumento) que no está directamente relacionada con el resultado potencial, pero sí

con la probabilidad de ser tratado. Los métodos de Variables Instrumentales utilizan esta idea para identificar el efecto del tratamiento para una determinada subpoblación de individuos.

El problema de todos estos estudios no experimentales es que requieren hacer uso de modelos econométricos, tanto para caracterizar la participación en el tratamiento como para estimar el efecto del mismo. Si alguno de estos modelos está incorrectamente especificado, el estimador del efecto del tratamiento será inconsistente. En las siguientes secciones desarrollaremos en más detalle cada uno de estos métodos.

II.2 Experimentos aleatorizados

En un experimento aleatorizado, el mecanismo que se utiliza para asignar a los individuos al grupo de tratamiento y al grupo de control es aleatorio. La aleatorización permite evitar el sesgo de selección endógena. Dado que el mecanismo de asignación es aleatorio, los resultados potenciales son independientes del mecanismo de asignación, porque la aleatorización en la asignación garantiza que las características de tratados y no tratados son virtualmente idénticas. Esto significa que, dado que el tratamiento es independiente de los resultados potenciales, se cumplirá que

$$E[Y_0 | D=1] = E[Y_0 | D=0],$$

por lo que ya no existe sesgo de selección, siendo posible calcular el efecto medio del tratamiento comparando las medias de los tratados y de los no tratados. Es importante destacar que, aunque nuestra discusión se centra en el cálculo de efectos medios, la aleatorización también permite identificar toda la distribución marginal de los resultados potenciales. Ello permitiría comparar el efecto del tratamiento no sólo en la media, sino también en cualquier otra característica de la distribución.

Para estimar el efecto del tratamiento, dado que la asignación al mismo es aleatoria, únicamente hay que calcular medias muestrales, análogas a las medias poblacionales de interés. Así por ejemplo, se puede utilizar la diferencia muestral media entre los resultados de individuos tratados ($\bar{Y}_1$) y no tratados ($\bar{Y}_0$):

$$\hat{\alpha} = \bar{Y}_1 - \bar{Y}_0,$$

Una cuestión relevante en este contexto es cómo realizar un contraste de diferencias de medias. Dado que el coste de los experimentos aleatorizados es habitualmente muy elevado, éstos se suelen realizar sobre muestras pequeñas, con lo que los contrastes basados en resultados asintóticos no son muy adecuados. Por ello,

se suele utilizar el llamado Contraste Exacto de Fisher, más adecuado para tamaños muestrales pequeños.

II.2.1 Contribuciones empíricas

La mayor parte de las contribuciones relevantes están centradas en la evaluación de la efectividad de programas de formación para individuos económicamente desfavorecidos, referidos además a EE.UU., en donde hay una gran tradición de programas piloto para la evaluación de políticas públicas desde principios de los años 70.

Entre los experimentos de mayor relevancia en economía, uno de los más antiguos es el de Impuesto Negativo sobre la Renta (NIT, *Negative Income Tax*). Estos experimentos se implementaron entre 1968 y 1982 por parte del gobierno federal de los EE.UU., y probablemente supusieron la investigación de políticas sociales públicas más ambiciosa hasta ese momento. Su objetivo principal era determinar los efectos sobre la oferta de trabajo de individuos desfavorecidos, cuya renta está por debajo del mínimo exento, de recibir un impuesto negativo sobre la renta, como alternativa de gasto social frente las transferencias ligadas a otros programas sociales. En el período señalado se realizaron cuatro experimentos distintos, cada uno de ellos muy localizado en determinadas áreas del país y dirigidas a colectivos desfavorecidos pero con distintas características. En particular, tan sólo uno de ellos se concentró en zonas rurales, y todos difieren en el grupo de tratamiento en características raciales, grado de pobreza y cuantía del impuesto negativo aplicado. Robins (1985) proporciona una panorámica de los principales resultados sobre la oferta de trabajo, medida bien en términos de horas anuales trabajadas, bien en términos de tasa de empleo, de los cuatro experimentos de NIT que se llevaron a cabo. Todos ellos muestran de forma consistente un efecto negativo sobre la oferta de trabajo, que se mantiene para los distintos colectivos considerados según raza (negros, hispanos, blancos) y otras características (edad, estado civil, sexo, etc.).

LaLonde (1986) explota datos de un experimento aleatorizado procedentes de un programa nacional de formación para trabajadores no cualificados desfavorecidos desarrollado en 1975, conocido como NSW. El seguimiento de tratados y no tratados dentro del programa se realiza en base a entrevistas regulares y permite evaluar la efectividad del programa comparando la evolución posterior al tratamiento de las ganancias salariales medias de tratados y no tratados. Además, LaLonde encuentra enormes diferencias al comparar los resultados de la evaluación con los que se obtienen mediante los métodos econométricos usuales cuando los datos de que se dispone no proceden de un experimento aleatorizado. Bloom y otros (1997)

evalúan la efectividad del programa nacional JTPA, aplicado en EE.UU. diez años después del programa NSW, para el caso de colectivos desfavorecidos de adultos y de jóvenes que han abandonado de forma temprana el sistema educativo. Para ello, cuentan con una gran muestra de datos de un experimento aleatorizado y evalúan el efecto sobre las ganancias del programa de formación. Al igual que en otros trabajos, encuentran que el impacto del programa es positivo pero de reducida magnitud en el caso de los adultos, y no significativo para los colectivos de jóvenes. Estos resultados están en línea con los obtenidos por Couch (1992) para datos experimentales del programa de formación NSW.

La comparación entre las ganancias, o entre las tasas de empleo o desempleo, etc., de tratados y no tratados a partir de datos experimentales permite estimar consistentemente el efecto medio del tratamiento, dado que el experimento aleatorizado garantiza que las características de tratados y no tratados son virtualmente idénticas. Sin embargo, no ocurre lo mismo si se pretenden comparar las historias laborales de tratados y no tratados con posterioridad al programa de formación, como en Ham y LaLonde (1996). La razón es que, aunque la asignación al programa sea aleatoria, la dinámica posterior de empleo y desempleo de individuos tratados se verá modificada como consecuencia del programa de formación, y diferirá por tanto de la dinámica para individuos no tratados. Por ello, aunque se disponga de datos experimentales, la evaluación del programa en términos de duraciones de empleo y desempleo precisa de un modelo estadístico formal que tenga en cuenta el problema de selección inducido por el propio programa de formación.

Otro marco de aplicación relevante se refiere a la evaluación de la incidencia de mayores recursos públicos en el rendimiento escolar. Cabe citar dos trabajos referidos a EE.UU. de Angrist y Lavy (1999) y Krueger (1999). Ambos trabajos se centran en el efecto del tamaño de la clase sobre el rendimiento, medido por la calificación obtenida en ciertos exámenes estandarizados.

En concreto, Krueger (1999) explota datos del experimento aleatorizado STAR (Student Teacher Achievement Ratio) realizado en el estado de Tennessee (EE.UU.) a cerca de 20000 estudiantes. Aunque la asignación inicial de los estudiantes por tamaño de clase es puramente aleatoria, la muestra disponible presenta, como en cualquier experimento, importantes desviaciones respecto al diseño ideal. En este sentido, este trabajo es una excelente ilustración de los problemas que surgen en este contexto por la pérdida de aleatoriedad y los procedimientos que se aplican para minimizar los sesgos resultantes. En particular, existe un problema de desgaste muestral que es claramente no aleatorio, dado que mientras la mayoría de los alumnos asignados a clases pequeñas tienden a continuar en éstas, algunos alumnos asignados a clases grandes abandonan el centro debido al descontento de sus padres. Además, se producen trasvases no aleatorios de alumnos que distorsionan

el tamaño real de las clases, comprometiendo los resultados del experimento. Por un lado, durante la duración del experimento tienen lugar trasvases de alumnos de clases grandes a pequeñas (a través de una realeatorización) debido a las quejas de sus padres. Por otro lado, algunos alumnos con problemas de comportamiento son trasvasados entre clases grandes y pequeñas.

El estudio muestra un mayor rendimiento de los alumnos asignados a clases de tamaño reducido, en contraposición a otros estudios previos basados en los mismos datos. Asimismo, se encuentra que el desgaste muestral, la re-aleatorización y los trasvases no aleatorios no afectan a los resultados principales. En contraposición, los estudios previos que explotaban estos datos realizaban generalmente una mera comparación de medias entre grupos sin tener en cuenta que el cambio temporal en la composición de los estudiantes distorsiona dicha comparación.

II.3 Estudios observacionales I: Matching y regresión

El principal problema de los estudios observacionales es que la selección de los individuos que van a ser sometidos al tratamiento no es aleatoria. Por tanto, los individuos tratados y no tratados pueden diferir en otras características, diferentes de la asignación o no al tratamiento. Si no se tienen en cuenta dichas características, se podría estar atribuyendo espuriamente al tratamiento un efecto sobre la variable de interés que realmente es debido a la influencia de otras características. Por tanto, para poder obtener efectos causales, es necesario controlar por las diferencias entre individuos tratados y no tratados.

A este respecto, cabe señalar que no siempre es posible controlar por todos los factores relevantes, por dos razones principales. En primer lugar, es posible que los datos disponibles no contengan algunas variables que determinan que el individuo sea elegible para el programa. En segundo lugar, en ausencia de un experimento aleatorizado es posible que las fuentes de datos de los individuos tratados y de los individuos no tratados que desean utilizarse como grupo de control sean distintas, afectando a la definición de las variables que reflejan las diferencias entre individuos.

II.3.1 Identificación: Selección en base a observables

En ausencia de datos experimentales, no se suele cumplir la independencia entre los resultados potenciales (Y_{1i} , Y_{0i}) y la variable que indica si el individuo ha sido sometido o no al tratamiento, D_i . Decimos que hay selección en base a observables si existe un vector de variables X tal que el tratamiento es independiente de los resul-

tados potenciales condicionando en dichas variables. En este caso, de forma análoga a lo que ocurría en el caso de un experimento aleatorio, por independencia condicional entre el tratamiento y los resultados potenciales tendremos que:

$$E[Y_1 - Y_0 \mid X] = E[Y_1 - Y_0 \mid X, D=1] = E[Y \mid X, D=1] - E[Y \mid X, D=0]$$

Bajo selección en observables, es posible calcular también los efectos del tratamiento sin condicionar en dichas variables. Para ello, debemos calcular los efectos de interés integrando sobre la distribución de dichas variables observables. Por tanto, tenemos que el efecto medio de interés definido previamente tiene la siguiente expresión:

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

Igualmente, condicionando en la asignación al tratamiento, tenemos que el efecto medio del tratamiento para los individuos tratados es:

$$\alpha_{SATE} = E[Y_1 - Y_0 \mid D=1] = \int (E[Y \mid X, D=1] - E[Y \mid X, D=0]) dP(X \mid D=1).$$

II.3.2 Estimadores de matching: Matching en variables observables

Cuando es un vector X de una única variable y toma un número pequeño de valores discretos, es sencillo construir estimadores de los efectos medios del tratamiento basados en las ecuaciones (matchte) y (sate). Simplemente, habría que calcular medias muestrales por celdas definidas en términos de los valores que tome la variable X . Supongamos que X toma J valores diferentes $\{X^1, \dots, X^j, \dots, X^J\}$. Denotemos por n^j el número de observaciones en la celda j . Sea n_1^j el número de observaciones correspondientes a individuos sometidos al tratamiento en la celda j , y sea n_0^j el número de observaciones correspondientes a individuos no tratados en la celda j . Finalmente, sea $\bar{Y}_1^j$ el resultado medio para los individuos tratados en la celda j e $\bar{Y}_0^j$ el resultado medio para los individuos no tratados en la celda j . Entonces, podemos calcular los efectos de interés de la siguiente forma:

$$\begin{aligned} \hat{\alpha}_{ATE} &= \sum_{j=1}^J (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n^j}{n}\right) \\ \hat{\alpha}_{SATE} &= \sum_{j=1}^J (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n_1^j}{n_1}\right) \end{aligned}$$

Sin embargo, cuando X es una variable continua, la estrategia de estimación es algo más complicada. Una posibilidad consiste en dividir el soporte de X en un

número finito de celdas y aplicar el mismo método que aplicaríamos si X fuera discreta. Este procedimiento se conoce como subclasificación.

Otra posibilidad es utilizar el llamado método de matching o emparejamiento. Dicho método consiste en emparejar el resultado de cada individuo tratado con el resultado de un individuo no tratado que tenga valores de la variable cercanos a los valores que dicha variable toma para el individuo sometido al tratamiento. Por ejemplo, cuando X es continua podemos estimar α_{SATE} como:

$$\hat{\alpha}_{SATE} = \frac{1}{n_1} \sum_{D=1} Y_i - Y_{m(i)},$$

donde $Y_{m(i)}$ es el resultado de un individuo no tratado que tiene valores de las variables cercanos a los de un individuo tratado, $X_i \approx X_{m(i)}$.

Ambos métodos son sencillos de aplicar cuando X contiene sólo una o dos variables. Sin embargo, si la dimensionalidad del vector X es grande, estos métodos no se pueden aplicar. En esos casos, la subclasificación podría dar lugar a celdas de X sin observaciones y, de igual forma, puede ser difícil encontrar para cada individuo tratado un individuo no tratado con valores de X similares. Como se verá más adelante, el método del *propensity score* trata de reducir la dimensionalidad de las variables contenidas en X a una única dimensión.

Dehejia y Wahba (2002) discuten tres cuestiones que afectan a la implementación del matching: si hacer o no matching con reemplazamiento de las observaciones del grupo de comparación, cuántas observaciones utilizar para emparejar cada individuo tratado, y qué procedimiento concreto de matching utilizar. En la disyuntiva entre hacer o no reemplazamiento de las observaciones del grupo de comparación, hay un compromiso entre la minimización del sesgo que se consigue con el reemplazamiento y la precisión de las estimaciones. En todo caso, al no reemplazar las observaciones, además del sesgo, surge el problema de que el orden en que se emparejan los individuos tratados no es trivial de cara a los resultados finales de la evaluación del tratamiento. La elección del número de observaciones que se emparejan con cada individuo tratado tiene también implicaciones en términos de sesgo (que se minimiza al emparejar con una sola observación) frente a precisión (que mejora aumentando el número de observaciones de comparación para cada tratado). Esta elección afecta de alguna manera al método concreto de matching a utilizar. En general, la respuesta a estas cuestiones depende de las características de los datos, y especialmente del grado de solapamiento entre tratados y no tratados de acuerdo con el *propensity score*. Si dicho solapamiento es muy limitado, será difícil conseguir un emparejamiento razonable sin reemplazamiento, y un matching con reemplazamiento sería en tal caso más recomendable. En cualquier caso, el rango para el que puede esti-

marse el efecto del tratamiento estará siempre sujeto a aquellos valores de los *propensity scores* para los que existan individuos no tratados comparables con los tratados.

II.3.3 Métodos de *propensity score*

Bajo selección en observables, llamamos propensity score a la probabilidad de recibir tratamiento, condicional a ciertas variables observables X tales que, condicionando en ellas, el tratamiento es independiente de los resultados potenciales, es decir,

$$P(X) = \Pr(D=1 | X).$$

Bajo ciertas condiciones (véanse Rosenbaum y Rubin, 1983 y Heckman, Ichimura y Todd, 1997), se puede demostrar que condicionar en el *propensity score* es suficiente para tener independencia entre el indicador del tratamiento y los resultados potenciales. Este resultado sugiere aplicar un procedimiento en dos etapas para estimar los efectos causales bajo selección en observables. En una primera etapa se estima el *propensity score* que caracteriza la probabilidad de participar en el programa, mientras que en una segunda se realiza el matching o la subclasificación en función de los valores del *propensity score*. Una de las principales ventajas de estos métodos es que la comparación de los individuos del grupo de no tratados respecto a los tratados se basa en la distribución conjunta de todas las variables observables, en vez de en la distribución marginal de cada variable observable por separado. Esto evita el riesgo de descartar observaciones del grupo de comparación por el mero hecho de que difieran respecto al grupo de tratados en el valor de una única variable.

Dehejia y Wahba (1999, 2002) muestran las ventajas del *propensity score* para determinar qué individuos del grupo de no tratados son relevantes para comparar con los tratados, permitiendo acometer de forma más sencilla el problema de evaluación. El *propensity score* puede estimarse por medio de un modelo *logit* u otro modelo de elección binaria. Además, el carácter restrictivo vinculado a la necesidad de parametrizar el *propensity score* se resuelve en gran medida gracias a la posibilidad de introducir potencias e interacciones de las variables observables. En este sentido, la sensibilidad de los resultados a la parametrización es mucho menor en el caso del *propensity score* que si optamos por parametrizar la ecuación (2). La especificación del modelo de elección binaria descansa en la independencia entre las variables que afectan al *propensity score* y la asignación o no al tratamiento, de manera que la distribución de dichas variables debe ser la misma para los individuos tratados que para los individuos del grupo de comparación. Así, para determinar la forma

funcional en que aparecen las variables en el *propensity score*, se parte de una especificación funcional para las variables (generalmente, la lineal), se estima el modelo de elección binaria, se dividen los valores ajustados del *propensity score* en estratos, y se contrasta si existen diferencias entre los grupos de tratamiento y de control dentro de cada estrato mediante contrastes de igualdad de distribuciones. En caso de no existir diferencias significativas, se acepta la especificación. De lo contrario, se añaden potencias de orden superior en las variables e interacciones entre éstas hasta que no se rechaza que las distribuciones entre ambos grupos son iguales. Una vez determinada la especificación, las observaciones se ordenan en orden ascendente de acuerdo con el valor estimado del *propensity score*, descartándose aquellas observaciones de no tratados con valores estimados del *propensity score* demasiado extremos. Con las observaciones que quedan, se procede a emparejar (con reemplazamiento) a cada individuo tratado con aquel individuo no tratado cuyo valor estimado del *propensity score* esté más próximo, descartándose las observaciones de no tratados que queden sin emparejar.

Por supuesto, como Dehejia y Wahba (1999) mencionan, es posible que la calidad de los datos del grupo de comparación limite enormemente las posibilidades de estimar el efecto del tratamiento. En particular, es posible que sólo unas pocas observaciones del grupo de comparación sean relevantes a pesar de que el tamaño de este grupo sea grande. Además, es posible que algunas variables importantes para determinar la asignación al tratamiento no se observen, en cuyo caso los métodos de *propensity score* no son útiles para estimar el efecto del tratamiento. Sin embargo, incluso en estas situaciones, estos métodos permiten establecer tanto la calidad del grupo de comparación como las posibilidades de estimar el efecto del tratamiento. En este sentido, el *propensity score* estimado puede servir como mínimo como un estadístico descriptivo del grado de coherencia entre el grupo de tratados y el grupo de comparación.

II.3.4 Métodos de regresión

Es posible interpretar la evaluación de políticas cuando existe selección en base a observables en términos de un modelo de regresión. Una regresión de Y sobre D y sobre X tendrá una interpretación causal si el tratamiento es exógeno una vez que se condiciona en las variables X . En este caso, una regresión por Mínimos Cuadrados Ordinarios (MCO) proporciona una buena aproximación a la función de respuesta causal media.

En concreto, si especificamos la esperanza condicional de Y dado D y X como una regresión lineal de Y sobre D , y X la interacción entre ambas, $D \times X$, tenemos

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

y

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

Por tanto, parámetros de interés como el efecto medio del tratamiento y el efecto medio seleccionado del tratamiento pueden estimarse sencillamente utilizando la regresión lineal:

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

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

Por supuesto, el estimador por MCO proporcionará una información pobre sobre los parámetros de interés en la medida en que la función $E(Y | D, X)$ no esté bien aproximada a través de una forma funcional lineal. En ese caso, se podrían estimar $E(Y | D=1, X)$ y $E(Y | D=0, X)$ mediante técnicas no paramétricas.

II.3.5 Contribuciones empíricas

LaLonde (1986) compara las ganancias de individuos que participan en un programa de formación con las ganancias de individuos no tratados pertenecientes a un grupo experimental de control por un lado y grupos de comparación no experimentales, por otro. El efecto estimado del tratamiento comparando datos experimentales del grupo de tratamiento y del grupo de control es sustancialmente diferente de la estimación de dicho efecto causal cuando el grupo de no tratados no está obtenido a partir de un experimento aleatorizado y utiliza técnicas de matching, entre otras. LaLonde concluye que es indispensable contar con datos de experimentos aleatorizados para identificar apropiadamente el efecto causal del programa de formación. Este resultado negativo acerca de los estudios observacionales es investigado por Dehejia y Wahba (1999), que encuentran enormes diferencias en la distribución de las variables observables del grupo de tratamiento y de los grupos de comparación no experimentales. Además, encuentran que las evaluaciones del programa de formación basadas en el propensity score se asemejan mucho más a las obtenidas con los datos experimentales.

Heckman, Ichimura y Todd (1997) aplican una extensión del método de matching que requiere condiciones más débiles para estimar consistentemente el efecto causal del programa nacional de formación JTPA, mencionado anteriormente, sobre las ganancias. Un aspecto central para resolver el problema de evaluación mediante procedimientos de matching es el *propensity score* $P(X)$. Los autores mues-

tran que, en contra de la propuesta de Ashenfelter y Card (1985) para un programa de formación similar, las variables asociadas a la búsqueda de empleo (fundamentalmente, las historias recientes de empleo y desempleo) son mejores predictores de la participación en programas de formación laboral que variables demográficas (edad, educación, raza) y que la historia previa de ganancias salariales.

II.4 Estudios observacionales II: Estimadores de diferencias-en-diferencias

En la sección anterior, hemos presentado cómo controlar por diferencias observables entre los individuos tratados y no tratados. Sin embargo, en la mayoría de los casos, existen buenas razones para pensar que ambos tipos de individuos también se diferencian en características inobservables que están correlacionadas con los resultados potenciales, incluso una vez que se controla por diferencias en características observables. En ese caso, los individuos tratados y no tratados dejan de ser directamente comparables. Este problema ha sido señalado, entre otros, por Heckman, Ichimura y Todd (1997).

La posible solución a este problema pasa por disponer de datos de la variable de interés antes y después de que el tratamiento se lleve a cabo. De esta forma, podemos comparar los resultados de los individuos tratados una vez recibido el tratamiento con los resultados de estos mismos individuos antes de recibir el tratamiento. El problema de esta comparación es que puede estar contaminada por tendencias temporales en la variable de interés o por el efecto de otro tipo de circunstancias que hayan ocurrido entre ambos periodos. No obstante, si existen individuos no tratados, éstos pueden utilizarse para identificar aquella variación temporal en la variable de interés que no es debida a la exposición al tratamiento. El estimador de diferencias-en-diferencias (DED) se basa en esta idea.

Igual que en las secciones anteriores, se define el efecto causal de un tratamiento en términos de la diferencia entre dos resultados potenciales. Pero ahora tenemos además una dimensión temporal en nuestros datos, por lo que los resultados potenciales también estarán indexados con un indicador de tiempo. Así, el efecto causal del tratamiento para el individuo i en el momento t quedará definido como $Y_{1i}(t) - Y_{0i}(t)$. De nuevo, el problema de identificación es que para cada individuo i y en cada periodo de tiempo t no observamos los dos resultados potenciales, por lo que no se puede calcular la anterior expresión para cada individuo. Por ello trataremos de estimar efectos medios del tratamiento.

El caso ideal para aplicar el método de DED es el que corresponde a una reforma no anticipada (es decir, un experimento natural) que afecta a ciertos subgru-

pos de la población objeto de estudio pero no al resto de la población. En tal caso, los individuos afectados son el grupo de tratamiento, mientras que el resto de individuos no afectados por la reforma constituyen el grupo de control. En el caso más simple, supongamos que hay dos períodos de tiempo y dos grupos, un grupo 1 de control compuesto por aquellos a los que no afecta el cambio de política y un grupo 0 de tratamiento (o experimental), que contiene a aquellos que sí se ven afectados.

Sean $\bar{Y}_1(0)$ e $\bar{Y}_1(1)$ las medias de Y para el grupo de control en los períodos anterior y posterior a la aplicación del tratamiento, e $\bar{Y}_0(0)$ e $\bar{Y}_0(1)$ los análogos para el grupo de tratamiento.

Si los datos se pudieran generar verdaderamente a partir de un experimento puro, entonces podríamos medir el efecto del tratamiento ignorando el primer período (previo al cambio de política) y comparando a los individuos del grupo de tratamiento con los del grupo de control una vez producido el cambio de política. Es decir, evaluaríamos el efecto del cambio de política calculando la diferencia de medias entre el grupo de tratamiento y el grupo de control después del cambio de política, $(\bar{Y}_0(1) - \bar{Y}_1(1))$. En la práctica, con datos no experimentales, este estimador tiene el problema de que parte de la diferencia entre la media del grupo de tratamiento y la media del grupo de control después del cambio de política puede deberse a diferencias sistemáticas e inobservables entre ambos grupos que no tienen nada que ver con el cambio de política.

Una medida interesante podría ser el efecto del cambio de política sobre el grupo de tratamiento, $(\bar{Y}_0(1) - \bar{Y}_0(0))$. El problema de esta medida es que la media del grupo de tratamiento puede cambiar a lo largo del tiempo (entre ambos períodos) por razones diferentes al cambio exógeno de política económica.

La medida apropiada para capturar el efecto del tratamiento (esto es, del cambio de política económica) consiste en comparar los cambios acaecidos en los grupos de tratamiento y de control, respectivamente, de manera que podamos controlar tanto las diferencias ex ante existentes entre ambos grupos como cambios acaecidos ajenos al cambio de política, es decir, $(\bar{Y}_0(1) - \bar{Y}_0(0)) - (\bar{Y}_1(1) - \bar{Y}_1(0))$.

El estimador de DED puede obtenerse también mediante una ecuación de regresión. Consideremos las variables binarias D , que toma el valor 1 si el individuo pertenece al grupo de tratamiento y 0 en otro caso, y S , que toma el valor 1 para el segundo período (después del cambio) y 0 para el primero (antes del cambio). En su versión más simple, la ecuación para analizar el impacto de un cambio de política es

$$Y = \beta_0 + \delta_0 S + \beta_1 D + \delta_1 (S \times D) + u,$$

El coeficiente de interés, δ_1 , está asociado a la interacción entre ambas variables binarias, D y S (cuyo producto no es más que una nueva variable binaria que es igual a 1 para los elementos del grupo de tratamiento en el segundo período). Es fácil comprobar que el estimador MCO de δ_1 se puede expresar como

$$\hat{\delta}_1 = (\bar{Y}_0(1) - \bar{Y}_0(0)) - (\bar{Y}_1(1) - \bar{Y}_1(0))$$

Por supuesto, para que el estimador DED evalúe consistentemente el efecto del cambio de política, es necesario que dicho cambio de política no esté sistemáticamente relacionado con otros factores inobservables (contenidos en u) que afecten a Y .

II.4.1 Contribuciones empíricas

Ashenfelter y Card (1985) evalúan el programa de formación laboral CETA, desarrollado en EE.UU. en 1976. Al no disponer de datos experimentales de tratados y no tratados, construyen un grupo de comparación y utilizan la información sobre salarios anterior al programa de formación para predecir las historias salariales de los tratados. Con ello, aplican un procedimiento de DED para evaluar el efecto del programa de formación comparando los salarios predichos de los tratados con los salarios observados de los no tratados, antes y después de recibir la formación. Para resolver el problema de no aleatoriedad de los individuos en cada grupo, se aplica un procedimiento en dos etapas en el que se corrige por la probabilidad de recibir formación basada en las historias salariales. Los autores encuentran una escasa robustez de la estimación del efecto causal a cambios en la especificación de la probabilidad de participación en el programa.

Card (1990) analiza el impacto en el mercado laboral de Miami, particularmente en el segmento de baja cualificación, de la entrada masiva de unos 120.000 inmigrantes cubanos en el verano de 1980. Dado el carácter no experimental de los datos, el autor establece una comparación de la evolución del mercado laboral antes y después de dicho suceso en otras cuatro ciudades estadounidenses que presentaban igualmente una proporción grande de negros e hispanos y un patrón de crecimiento económico similar a Miami en el momento del suceso. Los resultados sugieren que el efecto de esta llegada masiva fue relativamente pequeño, lo que puede deberse a que dicha llegada masiva previno la llegada de emigrantes domésticos que hubieran acudido en otro caso.

Card y Krueger (1994) analizan el impacto de un incremento en el salario mínimo legal en New Jersey, para lo que comparan el efecto sobre el empleo en restaurantes de comida rápida utilizando un estimador de DED que compara el empleo de establecimientos de dicha ciudad con el de establecimientos similares en

Pennsylvania antes y después de la reforma. Los resultados principales no muestran un efecto negativo significativo de dicha reforma sobre el empleo.

Meyer, Viscusi y Durbin (1995) evaluaron la importancia de los subsidios por enfermedad (que representan un porcentaje del salario hasta un tope máximo) en el tiempo de baja laboral aprovechando un experimento natural que tuvo lugar en Kentucky (EE.UU.). En julio de 1980, dicho Estado aumentó el tope máximo del subsidio por accidente o enfermedad laboral, lo que afectaba tan sólo a aquellos trabajadores con salarios altos que alcanzaban dicho tope. Este cambio de política disminuyó el coste de oportunidad de estar de baja laboral para los trabajadores de sueldos altos, pero no alteró las condiciones de aquellos cuyos salarios estaban por debajo del tope máximo. Esto permite implementar un estimador de DED para evaluar el efecto del subsidio por enfermedad, en el que además se puede controlar por la posible heterogeneidad individual utilizando características personales, características de la empresa y causas declaradas de la baja. Los principales resultados mostraron que el tiempo de baja para el colectivo de trabajadores con salarios altos aumentó como consecuencia de la reforma, mientras que se mantuvo invariante para los trabajadores de salarios más bajos que no se vieron afectados por la reforma. Puede por tanto concluirse que los subsidios por enfermedad tienen un importante efecto desfavorable en los incentivos de los trabajadores.

II.5 Estudios observacionales III: Variables instrumentales

En muchas ocasiones, los investigadores cuentan con alguna variable instrumental que ayuda a identificar el efecto causal de interés introduciendo selección exógena en el tratamiento. Dicha variable instrumental surge con frecuencia, en este contexto de evaluación de políticas públicas, de la existencia de un experimento natural en la que cierta variable relevante, al ser fijado su valor de forma administrativa, puede considerarse exógena.

En ese caso, los modelos de Variables Instrumentales (VI) proporcionan una estrategia de identificación alternativa. Estos métodos son particularmente adecuados para analizar experimentos sociales, ya que cuando las unidades experimentales son seres humanos, es difícil e incluso éticamente cuestionable realizar experimentos aleatorizados, o incluso forzar a que se cumplan los protocolos o las normas de los experimentos.

La estrategia de identificación en este caso es la siguiente. Supongamos que contamos con datos no experimentales y con información sobre características de los individuos, X , pero que condicionando en dichas características no se cumple la condición de independencia entre los resultados potenciales y el indicador del

tratamiento, D . Para simplificar, supongamos que tenemos en mente el siguiente modelo estadístico:

$$Y = \alpha + \beta D + \varepsilon$$

Si los datos fueran experimentales o procedieran de un experimento aleatorizado, la estimación por MCO de la anterior ecuación daría lugar a estimaciones consistentes del parámetro β y, por tanto, obtendríamos así el efecto causal del tratamiento. Sin embargo, si existe correlación entre la variable D y el término de error ε , o dicho de otra forma, si no se cumple la condición de independencia entre el resultado potencial y el efecto del tratamiento, la estimación por MCO dará lugar a estimaciones sesgadas del parámetro de interés. Una solución a este problema podría ser encontrar otra variable, Z , que sí satisface la anterior condición de independencia condicional y, además, está correlacionada con el tratamiento, D .

Suponiendo que el efecto causal sea homogéneo, es decir, el mismo para todos los individuos, es posible identificar el efecto del tratamiento como se hace tradicionalmente en modelos con variables explicativas endógenas. El parámetro de interés vendrá dado en la población por la siguiente expresión:

$$\beta = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)}$$

y una estimación consistente del mismo vendrá dada por el análogo muestral de la expresión anterior. En el caso en el que sea una variable binaria, se puede computar fácilmente como:

$$\hat{\beta}_{\text{WALD}} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0}$$

donde $\bar{Y}_1$ y $\bar{D}_1$ son las medias de Y y D para $Z=1$, mientras que $\bar{Y}_0$ y $\bar{D}_0$ son las correspondientes medias para $Z=0$. La expresión anterior se conoce en la literatura como Estimador de Wald.

Si por el contrario el efecto del tratamiento es heterogéneo entre individuos, la disponibilidad de variables instrumentales no es suficiente para identificar el efecto causal. Un supuesto adicional que ayuda a identificar dicho efecto es el supuesto de monotonicidad: cualquier individuo que estuviera dispuesto a ser tratado si fuera asignado al grupo de control, también debería someterse al tratamiento en caso de ser asignado al grupo de tratamiento. Bajo dichas condiciones, se puede identificar el llamado efecto medio local del tratamiento (LATE), que es el efecto medio para aquellos individuos cuyo valor D de cambiaría si lo hiciese el valor de Z (véase Angrist, 1990).

II.5.1 Contribuciones empíricas

Angrist (1990) evalúa el efecto causal sobre las ganancias salariales de haber prestado el servicio militar en el ejército de los EE.UU. durante la guerra de Vietnam. El problema aquí radica en que el hecho de ser veterano del ejército en dicho periodo es endógeno, al estar potencialmente correlacionado con componentes no observables de la ecuación de salarios, como la capacidad y otros. Sin embargo, es posible cuantificar el efecto causal utilizando como instrumento el suceso exógeno vinculado al alistamiento obligatorio por sorteo. Angrist calcula estimadores de Wald para ajustar las diferencias salariales de acuerdo con el resultado del sorteo debido a que algunos individuos con alistamiento obligatorio por sorteo no acudieron a filas, y a que otros no sujetos a alistamiento obligatorio se alistaron voluntariamente. Los resultados sugieren que el salario de los veteranos es en promedio un 15% menor que el de los no veteranos.

Angrist y Lavy (1999) evalúan el efecto causal del número de alumnos por clase en el rendimiento escolar en los colegios públicos de Israel. Para identificar dicho efecto, aplican un procedimiento de variables instrumentales utilizando las reglas administrativas exógenas que determinan el número máximo de alumnos en los centros. Los resultados indican que la reducción del número de alumnos por clase tiene un efecto sustancial en el rendimiento.

Edin, Fredriksson y Åslund (2003) proporcionan una interesante contribución acerca del efecto causal sobre las ganancias de los inmigrantes en Suecia del hecho de que vivan concentrados en enclaves étnicos. Un cambio en la política de inmigración sueca para solicitantes de asilo, que pasó a determinar exógenamente la asignación del lugar de residencia inicial de estos inmigrantes, proporciona un experimento natural. Dicho experimento natural permite utilizar la residencia asignada por parte de las autoridades como instrumento para resolver el problema de endogeneidad que supone la elección residencial de los individuos (que en un 48% cambian luego de residencia). Al utilizar la asignación administrativa del lugar de residencia para controlar por la endogeneidad de la elección residencial observada, se encuentra que la concentración étnica aumenta las ganancias de los inmigrantes menos cualificados alrededor de un 13%.

II.6 Otros enfoques

II.6.1 Efectos del tratamiento sobre distribuciones

Aunque la mayoría de los estudios existentes se centran en estimar diferencias en resultados medios, muchos investigadores se han interesado también en estudiar métodos para estimar el efecto de un tratamiento sobre toda la distribución de resultados. Estimar dichos efectos sobre distribuciones es relativamente sencillo utilizando datos experimentales, ya que únicamente se necesita comparar las distribuciones de los resultados potenciales. Para estimar efectos sobre toda la distribución cuando se dispone de datos observacionales se pueden utilizar métodos de variables instrumentales.

En cualquier caso, es posible estudiar los efectos del tratamiento sobre diferentes cuantiles de la distribución del resultado de interés utilizando técnicas de regresión sobre cuantiles (véase Koenker y Bassett, 1978). Supongamos que el cuantil de la distribución de dado y un conjunto de variables es lineal:

$$Q_{\theta}(Y | D, X) = \alpha_{\theta} \cdot D + X\beta_{\theta}$$

Cuando el tratamiento es aleatorizado o hay selección en base a observables, es el efecto del tratamiento sobre el cuantil. También es posible adaptar el modelo de regresión cuantílica al caso de variables instrumentales (véase Abadie, Angrist and Imbens, 1999).

II.6.2 Modelos de regresión discontinua

En algunas ocasiones, la asignación al tratamiento se determina de acuerdo con que las características del individuo excedan o no un cierto valor. Por ejemplo, la concesión de becas a estudiantes se suele hacer utilizando información sobre calificaciones previas de los mismos, que obviamente están correlacionadas con el resultado de interés (que pueden ser las calificaciones futuras). Por ello, de nuevo la simple comparación de resultados entre individuos tratados y no tratados no permite identificar adecuadamente los efectos causales del tratamiento. Sin embargo, si la correlación entre este tipo de variable y el resultado de interés es suave, se puede utilizar la discontinuidad creada por el tratamiento en el punto de corte para estimar el efecto del tratamiento en el resultado.

Consideremos el siguiente ejemplo. Supongamos que el tratamiento, D , consiste en la concesión de una beca en el momento de la admisión a la universidad siempre que la nota media de selectividad, X , esté por encima de cierto valor, Nuestro

interés se centra en estudiar el efecto de dichas becas sobre las calificaciones obtenidas en la carrera. Sean Y_1 e Y_0 las calificaciones potenciales si el individuo recibe y no recibe la beca, respectivamente. Como siempre, no observamos ambos resultados potenciales para ambos individuos, sino que únicamente se observa el resultado Y , que es igual a Y_1 en caso de haber recibido beca, e igual a Y_0 en caso contrario. Por supuesto, ambos resultados potenciales, Y_1 e Y_0 están correlacionados con X : en media, los estudiantes con mejores calificaciones en selectividad, son también los que obtienen mejores resultados en la universidad.

Para simplificar, supongamos que la media de Y_0 dado X es lineal:

$$E(Y_0 | X) = \mu + \beta X$$

Asimismo, supongamos que el efecto medio del tratamiento no depende de X :

$$E(Y_1 - Y_0 | X) = \alpha$$

Por tanto, tendremos que:

$$E(Y_1 | X) = \alpha + E(Y_0 | X) = \alpha + \mu + \beta X$$

Puesto que D está determinado dado X (ya que será igual a 1 siempre que X sea mayor que $\bar{X}$ y será igual a 0 en caso contrario), tendremos que

$$\begin{aligned} E(Y | X, D) &= D \cdot E(Y_1 | X) + (1 - D) \cdot E(Y_0 | X) \\ &= \mu + \alpha D + \beta X \\ &= (\mu + \beta \bar{X}) + \alpha D + \beta (X - \bar{X}). \end{aligned}$$

Por tanto, para estimar el efecto del tratamiento α , bastaría con regresar Y sobre D y sobre X , siendo el efecto de interés el coeficiente α asociado a D . De forma equivalente, de acuerdo con la última línea de la expresión anterior, podríamos regresar Y sobre D y sobre $\bar{X}$ y $X - \bar{X}$.

II.6.3 Cotas no paramétricas

En caso de que no sea posible aplicar ninguno de los métodos anteriores, aún es posible obtener algún tipo de efecto causal de un tratamiento siempre que la variable de interés esté acotada, es decir, cuando tenga un límite superior e inferior. En ese caso, es posible también poner cotas al efecto del tratamiento.

Por ejemplo, supongamos que la variable dependiente de interés es binaria, y por tanto su cota inferior es 0 y su cota superior es 1. En este caso, sabemos que el efecto medio del tratamiento estará entre cero y uno, por lo que será posible esti-

mar las cotas. No obstante, el problema con las cotas es que en muchas ocasiones son demasiado anchas y, por tanto, poco informativas (aunque siempre es posible hacer algún tipo de supuesto para reducir los intervalos).

II.6.4 Enfoque estructural

Frente a la literatura descrita de efectos de tratamiento, centrada en la idea de comparar un grupo de tratamiento y un grupo de control, el enfoque estructural se centra en la formulación de un modelo de comportamiento económico, de manera que los parámetros estructurales de dicho modelo son primitivas de las funciones de comportamiento de los agentes económicos implicados (véase Heckman y Vytlačil, 2005). Para poder evaluar el impacto de políticas públicas en el marco de modelos estructurales, es necesario recurrir a supuestos de exogeneidad y de invarianza ante cambios de política de determinados parámetros estructurales, de manera que los experimentos contra factuales que pueden realizarse sean inmunes a la llamada "crítica de Lucas".

La estrategia de evaluación de cambios de política se basa en simular el efecto sobre la función de comportamiento formulada ante cambios exógenos en ciertos parámetros cuyos valores dependen del contexto institucional en el que los agentes toman sus decisiones. El gran atractivo de estos modelos, que consiste en que los parámetros tienen interpretación económica clara, bien como parámetros de las funciones de comportamiento de los agentes, bien como parámetros que determinan el marco institucional relevante. Sin embargo, éste es también su aspecto más criticable, en la medida en que su validez está sujeta al cumplimiento de las hipótesis de formas funcionales, exogeneidad, e invarianza ante cambios de política. Asimismo, la identificación de los parámetros de estos modelos requiere generalmente supuestos muy restrictivos.

II.6.5 Contribuciones empíricas

Uno de los primeros trabajos que basan la estrategia de identificación y estimación del efecto causal en el diseño de regresión discontinua es Angrist y Krueger (1991). A la hora de evaluar el rendimiento de los años completados de educación sobre las ganancias laborales, debe tenerse en cuenta la potencial endogeneidad de la educación debido a la omisión de variables correlacionadas con la educación, como la capacidad, así como el hecho de que en general no existen características observadas del individuo que sean instrumentos válidos. Sin embargo, existe una regla

en la mayoría de los países, asociada a la edad de escolaridad obligatoria. En particular, las leyes educativas en EE.UU. exigen permanecer en la escuela hasta cumplir una determinada edad (16 ó 17 años, dependiendo del Estado). La regla asociada a estas leyes permite explotar un experimento natural para estimar el impacto de la educación obligatoria, basándose en que aquellos nacidos en los primeros meses del año empiezan su escolarización a mayor edad que los nacidos en los últimos meses. Por ello, para aquellos que desean abandonar la escuela cuando alcanzan la edad límite de escolaridad obligatoria, los nacidos en el primer trimestre del año tendrán menos estudios que los nacidos en el último trimestre del año. Por tanto, es posible estimar el efecto causal de la educación obligatoria o, en otras palabras, si aquellos que se ven obligados a permanecer más tiempo en la escuela a consecuencia de la escolarización obligatoria ganan más debido a su mayor nivel de educación. Los estimadores basados en el trimestre de nacimiento muestran un efecto de la educación ligeramente mayor que los estimadores MCO convencionales.

Van der Klaauw (2002) evalúa el efecto de las ayudas de estudio concedidas por los centros universitarios en la elección de centro por parte de los estudiantes utilizando datos de solicitudes y admisiones de un centro universitario de EE.UU. Es evidente que la política de ayuda financiera está positivamente relacionada con el currículum del estudiante. Además, hay información que no es observable para el centro, como las alternativas efectivamente disponibles para el estudiante (tanto en otros centros universitarios como en el mercado de trabajo). Sin embargo, es posible hacer uso de la regla de pre-clasificación para asignar ayudas de estudio que aplica el centro, que establece niveles de ayuda (y por tanto, discontinuidades en la regla de determinación de becas) en función de la calificación obtenida en pruebas estándar de conocimientos. Los resultados basados en regresión discontinua muestran que la política de ayudas ofrecidas por el centro universitario es un instrumento efectivo para captar mejores estudiantes por los que compiten con otros centros.

La estrategia de Angrist y Pischke (1999), que utilizan un procedimiento de variables instrumentales basado en la regla exógena asociada al máximo número de alumnos por centro, está basado de hecho en el diseño de regresión discontinua. Dicha regla induce una discontinuidad en las reglas de asignación de tamaño de las clases. En concreto, en cualquier colegio público, ninguna clase de un determinado curso puede tener más de 40 alumnos. Ello supone por ejemplo que si en un colegio un curso tiene 41 alumnos, habría que crear dos clases, con un tamaño medio de 20.5 alumnos. En el caso de un colegio con un curso de 80 alumnos, habría dos clases, pero si el número de alumnos pasara a ser 81, habría tres clases con un promedio de 27 alumnos. Por tanto, la regla de tamaño genera una discontinui-

dad que genera enormes variaciones en el promedio de alumnos por clase. Por tanto, para evaluar el efecto sobre el rendimiento académico del tamaño de la clase, la regla de tamaño puede usarse como instrumento.

En relación con la estrategia de establecer cotas no paramétricas, Manski (1995) trata de acotar el efecto de tener una estructura familiar estable sobre la obtención del título de bachillerato entre estudiantes. En este caso, sería un indicador de graduación en el instituto, y sería un indicador de vivir en una familia donde el padre y la madre estaban presentes cuando el individuo tenía 14 años.

Hotz, Mullin y Sanders (1997) estudian las perspectivas relativas de futuras adolescentes que se quedan embarazadas. Dada la evidente endogeneidad (respecto a las variables asociadas a las perspectivas futuras) del suceso de que la adolescente esté embarazada, es necesario contar con un instrumento adecuado. Los autores proponen el aborto por causas naturales como instrumento, pero encuentran que dicha variable no es exógena en todos los casos y es por tanto un instrumento contaminado. La razón es que malos hábitos de salud como fumar y beber aumentan significativamente el riesgo. Esto supone que los abortos por causas naturales pueden clasificarse como no aleatorios (asociados a malos hábitos) o aleatorios. Para este último caso, es posible explotar la exogeneidad del instrumento y establecer cotas al efecto causal. El precio que se paga al recurrir a cotas es la pérdida de precisión, que se refleja en la falta de evidencia concluyente acerca del efecto del embarazo sobre la obtención del grado de secundaria. Sin embargo, un resultado robusto es que las adolescentes embarazadas tienden a trabajar en promedio más horas y obtienen mayores ganancias en su primera edad adulta, si bien en algunos casos los resultados no están en contradicción con los estimadores convencionales de mínimos cuadrados y de variables instrumentales.

Entre las aplicaciones del enfoque estructural, cabe citar algunas contribuciones para evaluar el efecto del impuesto negativo sobre la renta (NIT). Aunque los datos con que se evalúa el efecto del NIT son datos experimentales, el enfoque estructural viene motivado en este contexto por varias razones. En primer lugar, la diferencia de medias entre tratados y no tratados no puede extrapolarse a una población diferente (que por ejemplo, tenga una distribución de la renta distinta y una proporción de familias por debajo de la línea de pobreza diferente) o a programas de NIT alternativos con distintos tipos impositivos y subsidios. En segundo lugar, como ocurre en la mayoría de los experimentos, pero muy especialmente en este caso, es posible que algunos individuos tratados y no tratados se beneficien simultáneamente de otros programas públicos, lo que puede contaminar la evaluación del programa, particularmente si ésta se basa en una comparación de medias entre tratados y no tratados. En tercer lugar, la variable respuesta de interés (oferta de trabajo) está potencialmente sujeta a censura o truncamiento por razones de res-

tricciones mínimas y máximas en las horas de jornada laboral, de manera que la oferta de trabajo efectiva puede diferir de la oferta de trabajo ideal en ausencia de dichas restricciones. La idea principal es partir de un modelo de oferta de trabajo microfundamentado en el que las horas trabajadas son función del salario neto de impuestos y de otras rentas no salariales. No obstante, Burtless y Hausman (1978) señalan que dicho salario neto tiene un carácter no lineal, inducido por el hecho de que el salario después de impuestos depende de la oferta de horas trabajadas (y es por tanto endógeno). Por ello, la decisión de oferta de trabajo depende de todo el conjunto presupuestario a que se enfrenta el individuo, que se caracteriza por segmentos lineales con diversos puntos de esquina, y que es no convexo para familias sometidas al experimento.

Burtless y Hausman (1978) modelizan el conjunto presupuestario completo para cada individuo y, con objeto de imponer las mínimas restricciones sobre las elasticidades de oferta de trabajo, se apoyan en la teoría de la dualidad, que permite relacionar la oferta de horas con la función de utilidad indirecta subyacente sin necesidad de especificar la función de utilidad a priori. Para ello, necesitan especificar una función de oferta de trabajo suficientemente general e imponer las condiciones de Slutsky. Hecho esto, es posible recuperar los parámetros de la función de utilidad indirecta y por tanto disponer del mapa de preferencias del individuo con objeto de simular el efecto de programas alternativos dado dicho mapa de preferencias. Dichas simulaciones se basan en la invarianza de las preferencias ante cambios en las características del programa del NIT. Por otro lado, los autores deben tratar el hecho de que para individuos con valores idénticos de todas las variables observables se observa una oferta de trabajo diferente. Para reflejar estas diferencias, los autores deben introducir términos inobservables justificados por dos aspectos. En primer lugar, la posible desviación entre horas de trabajo deseadas y observadas puede deberse a errores de medida y, sobre todo, a variaciones inesperadas en las horas trabajadas por diversas razones (despidos involuntarios, horas extraordinarias o reducción de jornada forzosas, etc.). En segundo lugar, por diferencias en las preferencias que no vienen reflejadas por las variables observables, para lo que se permite un término constante distinto para cada individuo. Los resultados principales sugieren una elasticidad de las horas trabajadas respecto al salario neto insignificante (de manera que cambios marginales en el tipo del NIT no tienen efecto) y una elasticidad renta de las horas trabajadas de un 5% (de manera que la cuantía del subsidio sí tiene efecto). Moffitt (1979) aplica un modelo más sencillo a los mismos datos, abordando de manera diferente la no linealidad del conjunto presupuestario, utilizando el tipo medio relevante para cada individuo en vez del tipo marginal (no lineal), y obteniendo resultados similares.

II.7 Análisis coste-beneficio

Con independencia de los datos disponibles, y por tanto del procedimiento empleado para identificar el efecto causal, el análisis coste-beneficio tiene una gran importancia para los expertos de política económica. En concreto, los resultados del análisis deben contribuir a dilucidar si la política económica aplicada es útil o debe cambiarse. En este último caso, los resultados pueden sugerir diseños alternativos para nuevos programas. No obstante, el hecho de que el efecto causal sea positivo es una condición necesaria pero no suficiente para que un determinado programa sea recomendable.

Además de ello, un aspecto primordial para el experto en política económica es la rentabilidad social del programa. Generalmente, los resultados asociados al efecto causal proporcionan los beneficios (que pueden medirse a corto plazo y proyectarse a un más largo plazo), que idealmente deben confrontarse a los costes para concluir en qué medida el programa es socialmente rentable. Una de las primeras tareas es considerar el efecto causal neto, es decir, excluyendo todo aquel beneficio para los tratados, como los subsidios, que sea un coste para el resto de la sociedad.

En todo caso, no siempre es fácil evaluar los costes totales ni los beneficios totales de un determinado programa. Dicha tarea requiere a menudo supuestos muy fuertes y por ello discutibles. En consecuencia, no son muchos los trabajos que, además de medir el efecto causal, evalúan los beneficios netos de un determinado programa. A continuación nos centraremos en algunos de los escasos trabajos empíricos antes citados que evalúan dichos beneficios netos.

En el caso de Bloom y otros (1997), el coste del programa por individuo viene dado por el coste unitario de los servicios de formación ofrecidos. Además, se concluía que el principal beneficio esperado del programa de formación JTPA era un aumento de las ganancias de los tratados, si bien sólo el aumento salarial neto de subsidios es socialmente relevante. Sin embargo, la disponibilidad de datos se reduce a 30 meses, y por ello es difícil extrapolar el impacto posterior del programa en las ganancias. Los autores toman una posición conservadora, consistente en ver si el beneficio neto al final de los 30 meses de finalizada la formación es o no positivo (lo que sería una cota inferior al beneficio neto del programa). La conclusión es que dicho beneficio neto es positivo, aunque modesto para los adultos (lo que justificaría la continuidad del programa), pero negativo para jóvenes (lo que justificaría eliminar el programa para este colectivo o sustituirlo por otro alternativo).

En el caso de Krueger (1999) para evaluar el aumento de los recursos educativos conducentes a reducir el número de alumnos por clase en los cuatro primeros años de escolarización, el autor hipotetiza acerca de reducir el número de alum-

nos por clase en un tercio. El incremento de costes asociado es fácilmente medible, siendo aproximadamente de 7400 dólares por año. Sin embargo, es mucho más difícil cuantificar los beneficios. El autor hace el supuesto discutible de que la ganancia en el rendimiento académico se mantiene constante en el futuro y por tanto puede estimarse el efecto potencial sobre las ganancias futuras, de manera que para una tasa de descuento anual del 3%, los beneficios de dicha política serían de 9.603 dólares para los hombres y 7.851 para las mujeres, lo que supone un beneficio neto positivo.

El marco del enfoque estructural permite evaluar los efectos sobre el bienestar de una determinada política (o de una reforma de política económica) en el marco de modelos de equilibrio parcial o de equilibrio general. Obviamente, la validez del resultado del análisis está sujeta, como se ha mencionado anteriormente, a la validez de los supuestos del modelo estructural.

III METODOLOGÍA PARA EL DISEÑO E IMPLEMENTACIÓN DE ESTUDIOS DE EVALUACIÓN DE POLÍTICAS SOCIALES

Nuria Rodríguez Planas
Universidad Autónoma (Barcelona)

III.1 Introducción

Dadas las crecientes limitaciones en la disponibilidad de recursos públicos, y las tendencias a reconsiderar la eficiencia del gasto público en cuestiones sociales, es cada vez más importante una distribución más efectiva y eficiente de dichos recursos. Ahora bien, para mejorar las políticas públicas y su eficiencia es necesario analizar previamente los efectos de dichas políticas. Países anglosajones (EEUU, Canadá ó Inglaterra) y nórdicos (Suecia ó Dinamarca) han iniciado ya este proceso. Es sin duda este asunto uno de los que en el futuro inmediato demanda una mayor atención tanto en el conjunto del Estado español como en cada una de sus comunidades autónomas.

El objetivo de este capítulo es ofrecer recomendaciones técnicas para el establecimiento de un sistema de base de datos que puedan ser utilizadas para evaluar políticas públicas. Con ese fin, en este capítulo se establece qué tipo de datos se necesitan para dicha evaluación. Más en concreto, se discuten los problemas metodológicos más frecuentes y se presenta un plan de análisis para la evaluación de un programa público.

El apartado 'Problemas metodológicos frecuentes' presenta los retos más frecuentes con los que se enfrenta el investigador en cuatro momentos claves de la evaluación. Primero, se discuten los retos del investigador durante el desarrollo del