# TÉCNICAS MICROECONOMÉTRICAS PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

**ANTONIO MORENO-TORRES GÁLVEZ (*)**

Ingeniero Industrial del Estado

*«Si buscas resultados distintos,*
*no hagas siempre lo mismo»*

Albert Einstein

La mejora de la efectividad de las políticas públicas requiere el conocimiento de las relaciones de causalidad entre intervenciones y resultados imputables a las mismas. En este artículo se presentan desde un punto de vista intuitivo, aunque sin pérdida de rigor formal, las diferentes técnicas que constituyen el estado del arte actual de la Microeconomía Cuantitativa o Microeconometría, que es la rama de la economía empírica que, completando las aplicaciones econométricas tradicionales de tipo descriptivo o predictivo, tiene por objeto el estudio de estas relaciones causales a partir del análisis de datos microeconómicos o microdatos.

La importancia de la comprensión de las relaciones de causalidad se ha de entender en el contexto de una filosofía de políticas basadas en evidencias que, haciendo uso intensivo de herramientas objetivas tales como el «Análisis Coste-Beneficio» o la «Evaluación de Programas», busca entre otras cosas mejorar las políticas públicas mediante la identificación de las intervenciones –políticas o programas- que con el menor coste generen un mayor impacto en los objetivos/resultados perseguidos, lo que resulta de la mayor importancia en un contexto como el actual de escasez de recursos y sospecha generalizada hacia lo público (Moreno-Torres, 2012).

No siendo nuevo el interés por esta cuestión, los enfoques que se han seguido para su estudio han evolucionado en el tiempo, pudiendo distinguirse básicamente dos: uno de carácter estructural y otro de carácter experimentalista.

## ENFOQUES ESTRUCTURAL Y EXPERIMENTALISTA EN EL ESTUDIO DE LA CAUSALIDAD

Las técnicas de modelado estructural fueron desarrolladas en los años 40 del siglo XX por el Premio Nobel Trygve Haavelmo en el marco de sus trabajos en la Universidad de Chicago para la comisión Cowles de investigación económica. Partiendo de la teoría económica sobre la toma de decisiones por los agentes económicos –consumidores y empresas-, asumiendo la homogeneidad de su comportamiento –lo que se traduce en la existencia de un agente tipo o representativo-, y a partir de datos agregados, se estiman los parámetros estructurales de un modelo de equilibrio general de ecuaciones simultáneas. Se trata pues de un enfoque «Macro-Estructural» –o de «Macro-Econometría»- que es robusto desde el punto de vista de su representatividad –o «validez externa»–, al permitir simular *ex-ante* y durante su fase de diseño los efectos de una política o programa cuando esta se implementa en contextos diferentes.

Un enfoque alternativo de carácter «Micro-Estructural» –o de «Micro-Econometría»- es el propuesto por el tam-

bién Premio Nobel de la Universidad de Chicago James Heckman que, reconociendo lo forzado del paradigma marshalliano del agente representativo, captura fenómenos reales –como la heterogeneidad de los agentes económicos o las desviaciones de sus comportamientos con respecto a los referenciados por la teoría (1)– mediante el análisis de datos obtenidos a nivel microeconómico –microdatos– (Heckman, 2000).

La disponibilidad de tales datos en cantidad y calidad, ya sea procedentes de encuestas diseñadas *ad-hoc* o de registros administrativos, es un fenómeno relativamente reciente que ha sido propiciado por dos hechos: por un lado, y desde la Segunda Guerra Mundial, la voluntad política de los poderes públicos para la implementación de la misión estadística de las Administraciones Públicas mediante la dotación de los recursos institucionales, humanos y técnicos adecuados; y por otro lado, y más en nuestros días, el desarrollo tecnológico que ha proporcionado las capacidades computacionales necesarias para la recogida, el almacenamiento y el tratamiento masivo de datos.

Es también en la explotación de microdatos en la que se basa el enfoque experimentalista o de resultados potenciales originariamente propuesto por Neyman en 1923 y formalizado por Rubin a partir de los años 70 del siglo XX (Holland y Rubin, 1988). En este caso, la identificación de relaciones causales se realiza a través de la emulación del ideal que constituye el experimento controlado aleatorio (*Randomized Control Trial-RCT*) del tipo del ensayo clínico comúnmente utilizado en medicina o psicología y en el que, al asignar un tratamiento administrativa y aleatoriamente dentro de una población que queda pues dividida en dos grupos de agentes estadísticamente similares –uno, grupo de tratamiento, que engloba a los agentes a tratar; y otro, grupo de control o de comparación, que lo hace con los que no–, se facilita la medición del impacto de aquel en forma sintética o reducida con un parámetro de tratamiento que se calcula como simple diferencia entre los resultados observados para ambos grupos.

En este artículo se describirán someramente los métodos observacionales inspirados en el enfoque experimentalista (2) que conforman el estado del arte actual de la denominada «Evaluación de Programas», si bien su investigación de frontera (3) se refiere a los modelos micro-estructurales que son necesariamente más complejos en tanto en cuanto la estimación de parámetros estructurales exige aprender mucho más –sobre comportamientos, funciones de utilidad, efectos sustitución y renta,…– a partir de unos mismos datos que son los que también se utilizan para la estimación de parámetros de tratamiento en los métodos observacionales.

Estos últimos presentan la ventaja de concentrarse en los datos en una aproximación de tipo ″caja negra″ que no exige en exceso de Teoría Económica - en palabras de Holland, se trataría de entender los efectos de las causas, más que las causas de los efectos–, pero que en contrapartida requiere solventar tres problemas fundamentales que se discuten en el siguiente epígrafe: uno que por ser de tipo conceptual es fácilmente salvable y dos para los cuales el ingenio estadístico-econométrico aporta soluciones bajo supuestos más o menos fuertes.

## ERRORES Y DIFICULTADES EN LA EVALUACIÓN DE POLÍTICAS. SOLUCIONES

### Correlación no es lo mismo que causalidad. Resultado no es lo mismo que impacto

Así, si con la notación $X->Y$ se expresa una relación de causalidad por la que la variable $X$ tiene un efecto en la variable $Y$, un primer problema es la «confusión entre correlación y causalidad», ya que la primera no implica necesariamente la segunda, sino que puede deberse a otros fenómenos tales como «causalidad inversa» (es la variable $Y$ la que realmente afecta a la variable $X$ y no al contrario: $Y->X$ según el convenio notacional fijado), «simultaneidad» ($X->Y$ y a la vez $Y->X$), la existencia de «variables ocultas» (hay una tercera variable $Z$ que causal y simultáneamente afecta a $X$ e $Y$, de manera que $Z->X$ y $Z->Y$), «errores de medición» (se observa $X$ en lugar de la verdadera variable causal $X^*$: $X^*->Y$) o «truncamiento de datos» (sólo se observan los valores de $Y$ correspondientes a ciertos valores de $X$).

Más en particular, si en un contexto de políticas públicas $X$ es una intervención e $Y$ la variable sobre la que se pretende actuar, una fuerte correlación positiva entre $X$ e $Y$, que se manifestaría en términos bayesianos en valores elevados de la probabilidad condicionada $P(X|Y)$, no garantiza de por sí la existencia de una relación de causalidad $X->Y$ como trata de ilustrar el siguiente ejemplo.

Sea $P(Y)$ la probabilidad de que una empresa dada sea innovadora, lo que vendría dado en gran parte por múltiples factores de tipo macro, de contexto o exógenos. Sea $P(X)$ la probabilidad de que una empresa tenga suscrito un convenio de colaboración y transferencia de tecnología con una universidad, que en este caso tendría carácter micro en el sentido de que puede actuarse sobre ella mediante políticas de fomento. En efecto, de la observación de valores altos de $P(X|Y)$, que sería la probabilidad de que una empresa innovadora tenga un acuerdo de este tipo, no se puede inferir de por sí su bondad –tampoco se pretende concluir lo contrario, tan sólo advertir de los riesgos de un excesivo optimismo- puesto que ello puede deberse simplemente a un alto valor de $P(X)$ o a que las empresas suscriptoras sean aquellas con una mejor dirección, lo que a su vez explicaría que fueran más innovadoras.

Desde el punto de vista de la toma de decisión sobre una política pública que fomentara estos convenios de colaboración, lo relevante es la probabilidad

RECUADRO 1
CONFUSIÓN ENTRE CORRELACIÓN Y CAUSALIDAD.
RESULTADOS E IMPACTOS

Todas las circunstancias que dan lugar a la confusión entre correlación y causalidad se traducen econométricamente en un único hecho formal, que es la correlación de $X_i$ con el término de error $\varepsilon_i$ cuando se expresa el vínculo entre $X$ e $Y$ con un modelo lineal del tipo $Y_i=\beta_0+\beta_1 \cdot X_i+\varepsilon_i$.

Equivalentemente, esto supone violar la «hipótesis de media condicional nula»: si con la notación habitual $E$ representa el operador esperanza matemática, no se cumple que $E(\varepsilon_i|X_i)=0$ para todo $X_i$, que es una de las condiciones que establece el «Teorema de Gauss-Markov» para que la estimación con un modelo convencional de regresión lineal de ajuste mínimo cuadrático ordinario sea insesgada, esto es, que $E(\widehat{\beta}_1)=\beta_1$. La dificultad estriba en que dicha condición para los residuos no es comprobable empíricamente, si bien por la propia construcción del modelo se cumple siempre mecánicamente para los residuos estimados.

En términos bayesianos, el error surge de la confusión entre las probabilidades condicionadas $P(X|Y)$ y $P(Y|X)$, lo que resulta equivalente a ignorar el ratio de probabilidades absolutas $P(Y)/P(X)$ ya que por el teorema de Bayes se tiene que

$$P(Y|X) \cdot P(X) = P(X|Y) \cdot P(Y)$$

y por tanto

$$P(Y|X)=\frac{P(Y)}{P(X)} \cdot P(X|Y)$$

El impacto probabilista sobre una variable de resultados $Y$ atribuible a la causa o intervención $X$ es por tanto la diferencia

$$P(Y|X) - P(Y|\neg X)=\frac{P(X|Y)\cdot P(Y)}{P(X)} - \frac{P(\neg X|Y)\cdot P(Y)}{P(\neg X)} = \frac{P(X|Y)\cdot P(Y)}{P(X)} - \frac{[1-P(X|Y)]\cdot P(Y)}{1-P(X)} = \frac{P(Y)\cdot[P(X|Y)-P(X)]}{P(X)\cdot[1-P(X)]}$$

FUENTE: Elaboración propia.

$P(Y|X)$ de ser una empresa innovadora habiendo suscrito uno y, más aún, el efecto adicional que en la innovación tuviera el formalizar un convenio con respecto a no hacerlo. Si con $\neg X$ se representa el evento complementario de $X$, este efecto incremental sería $P(Y|X)-P(Y|\neg X)$, diferencia que fijados $P(Y)$ y $P(X|Y)$ depende de la probabilidad *a priori* o tasa base $P(X)$ que es habitualmente ignorada en los análisis (recuadro 1).

Por tanto, este problema conceptual previo requiere para su evitación distinguir la parte de los resultados $Y$ que realmente sea atribuible a la intervención $X$, esto es, se ha de analizar esta en términos de su adicionalidad o efecto incremental sobre aquellos. Introduciendo ya el lenguaje y la notación que se utilizará en este artículo, denominando genéricamente agente –individuo o empresa, como en el ejemplo– a la unidad observacional a la que se le administrará o no un tratamiento –intervención– y si

$Y_{1i}$ es variable de resultados para el agente $i$-ésimo en el caso de recibir tratamiento

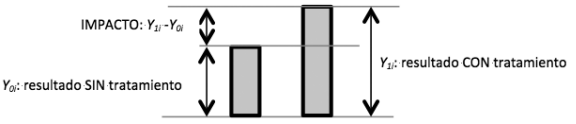
$Y_{0i}$ es variable de resultados para el agente $i$-ésimo en el caso de ausencia de tratamiento

entonces el impacto individual del tratamiento en el agente i-ésimo será la diferencia $Y_{1i}- Y_{0i}$ entre los resultados obtenidos para el agente en sus estados tratado y no tratado como se muestra en la figura 1.

Lamentablemente, este análisis incremental no se suele hacer a la vista de la muy común indistinción –intencional en muchas ocasiones– entre resultados e

**FIGURA 1**

**RESULTADOS E IMPACTOS**




FUENTE: Elaboración propia .

impactos en que se manifiesta la confusión correlación-causalidad.

Además, otros dos problemas complican la estimación de efectos causales mediante métodos observacionales: uno, el desconocimiento del denominado «contrafactual»– y otro, el fenómeno de selección.

## Problema fundamental de la inferencia causal. Soluciones científica y estadística

El primero de ellos, conocido como «problema fundamental de la evaluación» –o «problema fundamental de la inferencia causal»–, supone que no se puede calcular un efecto individualizado del tratamiento puesto que, a no ser que se disponga de un artefacto fantástico del tipo de una máquina del tiempo, no es posible observar a un agente someti-

do a tratamiento en su estado no tratado, y viceversa –el agente no sometido a tratamiento no es posible observarlo en su estado tratado–. Esto es, si $D_i$ es una variable binaria que toma el valor *1* o *0* en función de que se haya administrado o no tratamiento al agente *i*-ésimo, entonces las observaciones disponibles son $Y_{1i}$ si $D_i=1$ o $Y_{0i}$ si $D_i=0$, pero no los resultados potenciales que resultan relevantes para la evaluación, que son los contrafactuales respectivos $Y_{0i}$ *si* $D_i=1$ e $Y_{1i}$ *si* $D_i=0$ según se sintetiza con una tabla de doble entrada en el cuadro 1.

Nótese que la variable $D_i$ permite dividir la población de agentes en dos grupos, uno de agentes tratados caracterizado por $D_i=1$ y otro de agentes no tratados caracterizado por $D_i=0$ , derivándose la inobservancia del contrafactual del hecho de que un agente dado no pueda -salvo que tenga dotes mágicos- pertenecer simultáneamente a ambos grupos.

En el mundo de las ciencias físicas o naturales, en el que los comportamientos son predecibles de acuerdo a leyes o patrones estables, es posible una solución de carácter científico en la que como contrafactual se tome el resultado potencial que se hubiera observado en ausencia de tratamiento y con la ley de comportamiento operando invariantemente con su inercia. En el argot de la disciplina esto es lo que se conoce como «estimación pre-post» o «antes-después», en la cual no hay un grupo de control propiamente dicho, sino que se hacen observaciones antes y después –de ahí el nombre- del tratamiento para los agentes sometidos al mismo.

La «hipótesis de invariancia temporal» postula que sin tratamiento no debieran ser distintos los resultados en los dos momentos, por lo que toda diferencia en estos se ha de deber a aquel. Se tiene que

$$\begin{aligned} Pre\text{-}post &= Y_{1Después} - Y_{0Antes} = \\ &= (Y_{1Después} - Y_{0Antes}) + Y_{0Después} - Y_{0Después} = \\ &= (Y_{1Después} - Y_{0Después}) + (Y_{0Después} - Y_{0Antes}) = \\ &= Impacto + (Y_{0Después} - Y_{0Antes}) \end{aligned}$$

de manera que el estimador pre-post medirá el impacto cuando $Y_{0Después}=Y_{0Antes}$ , que es la expresión formal de la condición de inercia.

CUADRO 1
PROBLEMA FUNDAMENTAL DE LA INFERENCIA CAUSAL

| | $Y_{0i}$ | $Y_{1i}$ |
|---|---|---|
| $D_i=1$ | Contrafactual no observable | Observable |
| $D_i=0$ | Observable | Contrafactual no observable |

FUENTE:Elaboración propia.

Desafortunadamente, en el mundo de las ciencias sociales en el que se desenvuelven las políticas públicas esta solución científica no es en principio viable, al no existir leyes de evolución predecible. Si bien fenómenos cronológicos como la maduración en un ciclo de vida, la estacionalidad o el crecimiento en un ciclo macroeconómico pueden ser modelados para así corregir en la magnitud pertinente la estimación que el método pre-post proporciona en lo que se conoce como «estimación mediante serie temporal interrumpida» (4), los cambios de comportamiento de los agentes en un proceso de ajuste estratégico típicos de estos ámbitos son difíciles de anticipar y capturar en la estimación.

Por ello que, en estos casos, se adopta una solución de carácter estadístico que, en vez de estimar un efecto individual para cada agente, calcula en su lugar un efecto promedio –lo que equivale a imputar a los datos desconocidos con el valor promedio de los conocidos– para el conjunto de la población, dando lugar a parámetros como los del cuadro 2.

Si el *ATE* es relevante en caso de programas universales o de participación obligatoria, en aquellos programas focalizados o de participación voluntaria cabría estimar el *ATET* y el *ATEN*. Este último, aún sin tener una interpretación intuitiva clara –¿posible efecto de la incorporación al tratamiento de miembros del grupo de los no tratados?–, sí resulta de interés desde el punto de vista conceptual o analítico.

## Problema de selección. Selección endógena, autoselección y selección muestral

En cuanto al «problema de selección» –¿por qué unos agentes reciben tratamiento y otros no, aca-

CUADRO 2
SOLUCIÓN ESTADÍSTICA. PARÁMETROS DE TRATAMIENTO

| Parámetro de tratamiento | Expresión formal | Significado |
|---|---|---|
| ATE-Average Treatment Effect | $ATE=E(Y_{1i}-Y_{0i})=E(Y_{1i})-E(Y_{0i})$ | Impacto medio del tratamiento en el conjunto de la población |
| ATET-Average Treatment Effect on Treated o TT-Treatment effect on the Treated | $ATET=E[Y_{1i}-Y_{0i}\mid D_i=1]=E[Y_{1i}\mid D_i=1]-E[Y_{0i}\mid D_i=1]$ | Impacto medio del tratamiento en los tratados |
| ATEN-Average Treatment Effect on Nontreated o ATU-Average Treatment effect on Untreated | $ATEN=E[Y_{1i}-Y_{0i}\mid D_i=0]=E[Y_{1i}\mid D_i=0]-E[Y_{0i}\mid D_i=0]$ | Impacto medio del tratamiento en los no tratados |

Se tiene que $ATE=P(D_i=1)\cdot ATET+P(D_i=0)\cdot ATEN = P(D_i=1)\cdot ATET+(1-P(D_i=1))\cdot ATEN$

FUENTE: Elaboración propia.

RECUADRO 2
EL EXPERIMENTO SOCIAL EXPRESADO EN TÉRMINOS DE REGRESIÓN LINEAL

Si se ajusta un modelo de regresión del tipo $Y_i=\beta_0+\delta^*D_i+\varepsilon_i$ entonces el coeficiente en la variable binaria $D_i$ cuya estimación mínimo cuadrática es $\hat{\delta}=Cov(Y_i,D_i)/Var(D_i)$ refiriéndonos por *Cov* a la covarianza y por *Var* a la varianza, es precisamente el *ATE* puesto que $E(Y_i|D_i=0)=\hat{\beta}_0$ y $E(Y_i|D_i=1)=\hat{\beta}_0+\hat{\delta}$ de donde resulta $\hat{\delta}=E(Y_i|D_i=1)-E(Y_i|D_i=0)$. Dicha estimación es insesgada en el caso del experimento controlado aleatorizado en el que se asigna el tratamiento mediante una lotería, lo que constituye una guía para saber cuándo dar una interpretación causal al resultado de una regresión.

La violación de la «hipótesis de independencia de los resultados potenciales» en la selección que da lugar a estimaciones sesgadas del *ATE* se da cuando es $Cov(\varepsilon_i,D_i)\neq 0$ y simultáneamente también $Cov(\varepsilon_i,Y_i)\neq 0$. Equivalentemente, el término de error $\varepsilon_i$ estaría enmascarando una variable no observable *Z* que al omitirse en el modelo provoca un sesgo en la estimación cuyo signo se puede predecir: si *a* es la correlación entre la variable *Z* omitida y la variable *D* de participación, y *b* es la correlación entre Z y la variable *Y* de resultados, entonces el signo del sesgo es el del producto $a \cdot b$

La ventaja de la formulación del experimento en términos de regresión lineal es que se puede, por una parte, incorporar en su especificación variables de control $X_i$ por las diferencias observables entre los grupos de tratamiento y control, y por otra, tratar la heterogeneidad en la respuesta a través de un término de interacción $X_i \cdot D_i$ de manera que un modelo completo, en el que dependen de $X_i$ tanto las ordenadas en el origen como las pendientes, sería del tipo $Y_i=\alpha+\beta^*D_i+\lambda \cdot X_i+\gamma \cdot X_i \cdot D_i+\varepsilon_i$

Además, se puede precisar la significatividad estadística del ajuste en un contraste de una hipótesis base o nula que permita el establecimiento de un intervalo dentro del cual, con cierto nivel de confianza, no se podrá rechazar que esté el verdadero valor del parámetro a estimar. La potencia del contraste, que depende entre otros factores de la cantidad de datos, mide su poder explicativo entendiendo como tal su capacidad para rechazar la hipótesis base cuando una hipótesis alternativa dada es cierta.

FUENTE: Elaboración propia.

bando englobados en uno y otro grupo? o, equivalentemente, ¿por qué se observa lo que se observa en lugar de sus contrafactuales?–, este provoca que, una vez adoptada la solución estadística de calcular efectos promedio del tratamiento, la simple diferencia de las medias de los resultados observados para los respectivos grupos –tratados y no tratados– generalmente sea un cálculo erróneo que nos lleve a conclusiones inocentes como las ilustradas por la paradoja de Roy (1951). En efecto, se tiene que

$$E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]=$$
$$=E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]+E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=1]=$$
$$=\{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]\}+\{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]\}=$$
$$=ATET + Sesgo\ de\ Selección$$

es una estimación ingenua del *ATET* en presencia del fenómeno habitual de «selección endógena», que se da cuando la participación en un programa depende de manera consciente o inconsciente de factores relacionados con el resultado potencial (5), lo que formalmente supondría que $E[Y_{0i}|D_i=1]\neq E[Y_{0i}|D_i=0]$.

Esta selección puede ser o bien debida a factores observables (cuando por ejemplo es el propio administrador del programa quien marca la elegibilidad en base a variables demográficas o censales, lo que técnicamente se solventa de manera sencilla controlando por esos observables), o bien debida a factores inobservables (como son por ejemplo la motivación, la predisposición o la habilidad del tratado) en lo que constituye una «autoselección» de más complicada solución. Con selección exclusivamente basada en observables y en el caso de homogeneidad del tratamiento, los efectos *ATE*, *ATET* y *ATEN* resultan coincidentes, cosa que no ocurre en el caso de heterogeneidad en el cual los miembros de los dos grupos responden de manera diferente.

Un caso particular denominado «selección muestral» se da en situaciones con truncamiento de datos para los cuales Heckman (1976, 1979) propone un método bietápico (método *Heckit*) que usa una variable latente para estimar así la ley de decisión de los agentes con un modelo estructural de participación, cuyos residuos se incorporan en forma de ratio inverso de Mills como control en la estimación de efectos causales (corrección de Heckman), en una técnica denominada de «funciones de control».

## EL EXPERIMENTO SOCIAL COMO PARADIGMA, SÍ, PERO...

Es precisamente la evitación del sesgo de selección, pues por construcción los resultados son independientes del mecanismo de asignación de un agente a un grupo dado (tratamiento o control) al ser aquel de carácter aleatorio, lo que convierte al experimento social controlado –el *RCT* ya citado- en el paradigma de estimación causal para los métodos observacionales. Formalmente,

$$E[Y_{0i}|D_i=1]=E[Y_{0i}|D_i=0]=E(Y_{0i})$$
$$E[Y_{1i}|D_i=1]=E[Y_{1i}|D_i=0]=E(Y_{1i})$$

y consecuentemente se puede obtener el *ATE*, que por definición es igual a $E(Y_1)-E(Y_0)$, a partir de datos observados mediante la diferencia $E(Y_{1i}|D_i=1)-E(Y_{0i}|D_i=0)$ calculada directamente o a través de un modelo de regresión (recuadro 2). Es decir, se cumplen las condiciones para que el grupo de los no tratados –denomi-

RECUADRO 3
DESERCIÓN Y SUSTITUCIÓN EN EXPERIMENTOS SOCIALES. ESTIMADOR DE BLOOM Y ESTIMADOR DE WALD

Definiendo las siguientes variables binarias

$R_i$ que toma el valor 1 si el agente *i*-ésimo fue asignado al grupo de tratamiento y 0 si lo fue al grupo de control
$T_i$ que toma el valor 1 si el agente *i*-ésimo recibió tratamiento y 0 si no
$T^*_i$ que toma el valor 1 para aquellos agentes del grupo de control que hubieran sido tratados de haber estado en el grupo de tratamiento -por lo que es una variable latente no observable- y 0 si no
$S_i$ que toma el valor 1 si el agente *i*-ésimo del grupo de control recibió tratamiento sustitutivo y 0 si no

se tiene que sin deserción es $P(T_i=1,R_i=1)=1$ y sin sustitución es $P(S_i=1,R_i=0)=0$, y en el resto de circunstancias $ITT=E[Y_{1i}|R_i=1]-E[Y_{0i}|R_i=0]$ sería una estimación a la baja del *ATE*. Bajo ciertos supuestos, este podría aproximarse según se muestra en el siguiente cuadro en el que $Y_2$ es el resultado observado en caso de recibir tratamiento sustitutivo

| Desviaciones de la asignación | Hipótesis | Aproximación del ATE | Nombre |
|---|---|---|---|
| Deserción, no sustitución | $E[Y_i\|T_i=0,R_i=1]=E[Y_i\|T^*_i=0,R_i=0]$ (semejanza de los desertores) | $ITT/P[T_i=1,R_i=1)$ | Estimador de Bloom |
| Deserción y sustitución | $E[Y_{1i}-Y_{0i}\|T_i=1,R_i=1]=E[Y_{2i}-Y_{0i}\|S_i=1,R_i=0]$ (semejanza del tratamiento sustitutivo) | $ITT/(P(T_i=1,R_i=1)-P(S_i=1,R_i=0))$ | Estimador de Wald (Bloom generalizado) |

El sentido del estimador de Bloom resulta de considerar el total de los enrolados en el experimento divididos en dos grupos, tratados y desertores, lo que permite expresar el impacto por agente enrolado en el programa –lo que precisamente mide el *ITT*- como suma ponderada de los impactos para estos dos grupos tomando como pesos respectivos las proporciones de los mismos $P(T_i=1,R_i=1)$ y $P(T_i=0,R_i=1)$ de manera que, asumiendo que el programa no tenga impacto para los desertores –«hipótesis de semejanza de los desertores»: piénsese en el sesgo que introduce en la estimación su violación, como por ejemplo en el caso de un programa de escolarización en el que sólo permanecen los buenos alumnos-, se tendría que $ITT=P(T_i=1,R_i=1)\cdot ATE$, de donde se deduce la expresión que estima el *ATE* como un escalado del *ITT*.

En caso de tratamiento sustitutivo, y con la misma mecánica de entender el impacto como una suma ponderada, debiera minorarse el *ITT* en la contribución de aquel, que en la «hipótesis de semejanza del tratamiento sustitutivo» sería $P(S_i=1,R_i=0)\cdot ATE$, de manera que $ITT=P(T_i=1,R_i=1)\cdot ATE-P(S_i=1,R_i=0)\cdot ATE$, de donde se deduce la expresión del estimador de Wald para el *ATE*, también conocido como estimador de Bloom generalizado.

FUENTE: Elaboración propia.

nado grupo de control en este contexto– constituya un grupo de comparación adecuado.

Sin embargo, y aparte del hecho de que sólo permiten una evaluación *ex-ante*, los experimentos sociales pueden presentar una serie de problemas (Heckman y Smith, 1995):

**En su implementación práctica**, como en el caso de programas que respondan a derechos universales –en los que la asignación de agentes al grupo de control suscitaría consideraciones éticas–; o los excesivos costes que puede acarrear el propio experimento.

**En sus aspectos de «validez interna»** –¿son acertadas las inferencias para el contexto particular del programa?– como la aleatorización incorrecta o la existencia de deserciones y sustituciones, efectos placebo, alteración de comportamientos al sentirse observado –efecto *Hawthorne*–, sesgo del experimentador –que se traduce en una especial dedicación hacia el agente tratado, lo que se puede evitar con los experimentos«doblemente ciegos»–, o la medición incorrecta de resultados.

**En sus aspectos de «validez externa»** –¿son extrapolables las inferencias a otros agentes, contextos geográficos, institucionales o históricos, o variables de resultados?–, como es el caso de la no representatividad de las conclusiones de experimentos de pequeña escala –no es lo mismo replicar un programa que escalarlo–, o los efectos de equilibrio general –que incluyen contagios y otro tipos de impactos que se descartan con la «hipótesis de estabilidad del valor del tratamiento unitario» (*SUTVA*, acrónimo de *Stable Unit Treatment Value Assumption*) que asume también que el tratamiento para todos los agentes es comparable–.

Con respecto a la aleatorización, por las leyes de la estadística se dan ocasiones como las que Rubin llama del "doctor perfecto", que sería aquel que sólo trata a aquellos pacientes que sabe *a priori* que van a responder positivamente, lo que resultaría en una estimación sesgada al alza de la efectividad de sus actuaciones. Por ello es necesario comprobar que los grupos de tratamiento y control estén balanceados –son estadísticamente similares– antes de suministrar el tratamiento, lo que evidentemente sólo es posible hacer para atributos observables.

Los fenómenos de deserción –un agente del grupo de tratamiento no se presenta a recibirlo o abandona el mismo– y sustitución –un agente en el grupo de control se busca por su cuenta un tratamiento similar al del programa- conllevan reinterpretar la estimación con datos observados como un efecto medio de ofertar un programa o *ITT* (*Intent to Treat*) a partir del cual pue-

CUADRO 3
ESTRATEGIAS DE IDENTIFICACIÓN

| De menor a mayor sofisticación | En función de si la selección está basada en características observables o no observables | De mayor a menor información contenida en los datos sobre el mecanismo de asignación/proceso de selección |
|---|---|---|
| • Pre-post (solución científica)<br>• Comparación de cohortes<br>• Sección cruzada-corte transversal<br>• Diferencias en diferencias/Efectos fijos/ Métodos de panel<br>• Emparejamiento o *matching*<br>• Variables instrumentales<br>• Regresión discontinua<br>• Funciones de control | • Experimentos sociales-*RCT*<br>• Métodos observacionales no experimentales:<br>✓ Con selección basada en características observables<br>– Emparejamiento paramétrico: Sección cruzada-corte transversal<br>– Emparejamiento no paramétrico o *matching*<br>✓ Con selección basada en características no observables<br>– Variables instrumentales<br>– Diferencias en diferencias/Efectos fijos/Métodos de panel<br>– Regresión discontinua | • Asignación aleatoria «forzada»: Experimentos sociales-*RCT*<br>• Emparejamiento o *matching*<br>• Modelización del proceso de selección: Funciones de control<br>• Variables instrumentales<br>• Asignación aleatoria «casual»: Diferencias en diferencias/Efectos fijos/Métodos de panel<br>• Asignación determinística: Regresión discontinua |

FUENTE: Elaboración propia.

de aproximarse un valor del *ATE* bajo ciertas hipótesis que dan lugar a los estimadores de Bloom y Wald (recuadro 3). El *ITT* resulta valioso en el caso de programas de participación voluntaria en los que por definición los participantes se autoseleccionan.

Todos estos inconvenientes de los experimentos sociales esbozados motivan la búsqueda de alternativas, que de manera genérica se conocen como «estrategias de identificación», y que tienen por objetivo construir a partir de las observaciones disponibles un contrafactual razonable y/o considerar la autoselección en variables no observables con el fin último de solventar así el «problema de identificación»: asegurar que la relación entre un tratamiento y un resultado es causal. Su clasificación puede hacerse conforme a diferentes taxonomías como las recogidas en el cuadro 3 del cual la primera de ellas se utilizará como orden para la exposición en este artículo.

Es importante resaltar que los parámetros que resultan del ajuste econométrico de un conjunto de datos no informan de por sí de una relación causal salvo que sean leídos a la luz de un modelo de causalidad -estrategia de identificación- cuyas condiciones de plausibilidad se habrán de contrastar entendiendo –ya que no se puede hacer controlando- el proceso por el que los agentes acaban siendo tratados o no, ya sea por decisión de quien administra el programa o propia, lo que vendrá dado por un marco institucional y de comportamiento de los agentes concreto.

Esta dimensión contextual condiciona la especificidad de cada evaluación y fuerza al abandono de la quimera de la existencia de un método óptimo universal, lo que en general complica la disciplina y tiene como consecuencia práctica que la estrategia de evaluación de cada política haya de ser concebida caso por caso y en el mismo momento en el que se esté diseñando.

## ESTRATEGIAS SENCILLAS DE IDENTIFICACIÓN

### Comparación de cohortes o comparación con-sin (tratamiento)

La validez interna de un experimento social está en principio salvaguardada por la aleatorización que, al realizarse entre agentes estadísticamente similares, asegura que los impactos no puedan verse influidos por diferencias entre aquellos, sino que se deban en exclusiva al tratamiento. Sin aleatorización, en ciertas ocasiones pueden estimarse consistentemente efectos causales utilizando un «cuasi-grupo de control» o cohorte de agentes similares a los que reciben el tratamiento. Con un tamaño muestral adecuado se compensarían los efectos idiosincráticos en los agentes tratados, aunque aún existiría el riesgo de factores de selección endógena en su grupo frente a la cohorte de comparación.

### Sección cruzada

En una «sección cruzada» o «corte transversal» se explota una ″fotografía″ de datos correspondientes a un momento posterior al tratamiento, de manera que, si se asume que este es asignado de manera exógena –esto es, sin autoselección– entre grupos de tratamiento y control que son idénticos en términos de nivel de la variable de resultados, la diferencia que se observe en esta tras el tratamiento habrá de ser atribuida al mismo. La interpretación del modelo se alinea con la del experimento social si se consideran las observaciones de la sección cruzada como grupos de agentes sometidos a diferentes niveles de tratamiento y sin diferencias estadísticas entre ellos que puedan explicar parte de los impactos.

Es el más sencillo de los métodos paramétricos –aquellos que exigen la formulación de una relación funcio-

nal entre las variables del modelo- de estimación y permite controlar de manera simple por variables observables, aunque por el contrario adolece de inconvenientes como la no contemplación de heterogeneidades en el tratamiento –no confundir con heterogeneidades en la respuesta– su invalidez en caso de que los grupos de tratamiento y control presenten un nivel diferente en la variable de resultados, y aquellos otros derivados de la endogeneidad que pueda ser causada por autoselección en base a inobservables o cualquier otro de los motivos esbozados al tratar las fuentes de confusión entre correlación y causalidad. Lo fuerte de sus asunciones deja de corolario que como regla general se ha de ser escéptico sobre la consistencia de las estimaciones de sección cruzada.

## Experimentos naturales. Diferencias en diferencias

La circunstancia de que en ciertas ocasiones un hecho socioeconómico clasifique automáticamente a dos grupos preexistentes en uno de tratamiento y otro de control constituye una oportunidad inesperada -que no se puede planificar *a priori*- para hacer una evaluación *ex-post* o retrospectiva que solvente el problema de selección. Esto es lo que se conoce en la literatura sobre inferencia causal como «experimento natural» o «cuasi-experimento» –por utilizar un cuasi-grupo de control–, lo que da lugar a la estrategia de identificación de «diferencias en diferencias» (DD) o «dobles diferencias» que mejora a la sección cruzada al superar debilidades suyas como son las exigencias de un mismo nivel para la variable de resultados antes del tratamiento y, sobre todo, de exogeneidad del proceso de asignación, que sería en principio plausible en este caso, dado que la aleatorización es creada accidentalmente por fuerzas externas.

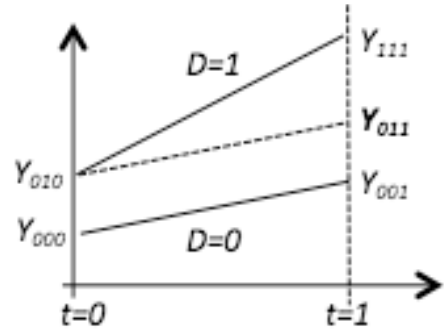
El método, que también admite formulación en términos de regresión lineal (recuadro 4), minora la diferencia entre resultados de los dos grupos después del tratamiento en la magnitud de la diferencia antes, para tener así en cuenta las disimilitudes entre ambos que pudieran estar afectando a los resultados, que por su parte habrán de presentar una tendencia común previa para que las estimaciones del *ATET* sean consistentes. En efecto, utilizando una notación de tres subíndices $Y_{JDt}$ en la que $J$ tendría el mismo significado que el subíndice principal para la variable de resultados $Y$ en el cuadro 1, se tiene que:

$$\begin{aligned} DD &= E[(Y_{111}-Y_{001})-(Y_{010}-Y_{000})] = \\ &= E[(Y_{111}-Y_{001})-(Y_{010}-Y_{000})+Y_{011}-Y_{011}] = \\ &= E(Y_{111}-Y_{011})+E[(Y_{011}-Y_{010})-(Y_{001}-Y_{000})] = \\ &= ATET+E[(Y_{011}-Y_{010})-(Y_{001}-Y_{000})] \end{aligned}$$

Es decir, $DD=ATET$ si se cumple que $E(Y_{011}-Y_{010})=E(Y_{001}-Y_{000})$ que es la expresión formal de la «condición de tendencia común» para ambos grupos en ausencia de tratamiento y que en la figura 2 se manifiesta con el paralelismo de la línea punteada –que representaría la evolución contrafactual del grupo de tratamiento en ausencia de este– con la línea continua de evolución del grupo de control.

**FIGURA 2**

**TENDENCIA COMÚN EN EXPERIMENTOS NATURALES**




FUENTE: Elaboración propia.

Algunos ejemplos ya clásicos de aplicación de *DD* son los discutidos por Card y Krueger (1994) sobre el impacto de los salarios mínimos en los niveles de empleo (6) o por Grogger (2002) sobre los efectos de la zonificación urbana en la criminalidad causada por el pandillismo (7). En ellos se ponen de manifiesto sus dificultades que se refieren fundamentalmente a la factibilidad del supuesto de tendencia común –cuya comprobación sólo se puede aproximar analizando series históricas con suficientes datos– y a la existencia de selección temporal idiosincrática –que afecta a sólo uno de los grupos- debida a cambios de comportamientos de los agentes y que compromete la consistencia de la estimación (Ashenfelter, 1978) (8).

Adicionalmente, persistirían los inconvenientes de la más sencilla de las estrategias de identificación con "película" de datos –necesariamente más costosa que una simple "fotografía"– correspondientes a momentos antes y después del tratamiento, que es la pre-post ya descrita basada en la solución científica para la lectura de una única serie temporal –dos serían las que se utilizan en el método *DD*- y que, aparte de lo exigente de la hipótesis de inercia, presenta complicaciones cuando son varias las intervenciones que coinciden en el tiempo –lo que dificulta el deslinde de los efectos atribuibles a cada una–, cuando hay efectos de tipo dinámico –por los cuales los eventuales efectos tratan en manifestarse– o cuando los impactos no son suficientemente grandes con respecto a los errores –relación señal/ruido pobre- en la extrapolación que al fin y al cabo subyace detrás de estas técnicas.

## Métodos de panel

Una mejora del método en términos de validez interna puede lograrse incorporando varios grupos de tra-

RECUADRO 4
EL EXPERIMENTO NATURAL FORMULADO EN TÉRMINOS DE REGRESIÓN LINEAL

La estrategia *DD* de diferencias en diferencias puede formalizarse como modelo de regresión, con las ventajas añadidas de poder realizar un contraste de hipótesis sobre los estimadores del modelo o incluir otras variables de control. Si definimos la variable binaria $I_j$ que toma los valores 0 y 1 para los instantes antes y después del tratamiento, respectivamente, se tiene que si formulamos el modelo $Y_{JDt}=\beta_0+\beta_1 \cdot I_j+\beta_2 \cdot D_i+\delta \cdot D_i \cdot I_j+\varepsilon_i$ el coeficiente $\delta$ del término de interacción $D_i \cdot I_j$ será el estimador *DD*. En efecto, se tiene que

| | **Antes ($I_j=0$)** | **Después ($I_j=1$)** |
|---|---|---|
| **Grupo de Tratamiento ($D_i=1$)** | $E(Y_{010})=\hat{\beta}_0+\hat{\beta}_2$ | $E(Y_{111})=\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\delta}$ |
| **Grupo de Control ($D_i=0$)** | $E(Y_{000})=\hat{\beta}_0$ | $E(Y_{001})=\hat{\beta}_0+\hat{\beta}_1$ |

Y por tanto:
$\hat{\beta}_0=E(Y_{000})$
$\hat{\beta}_1=E(Y_{001}-Y_{000})$ (diferencia después-antes entre los no tratados)
$\hat{\beta}_2=E(Y_{010}-Y_{000})$ (diferencia tratamiento-control antes)
$\hat{\delta}=E[Y_{111}-(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2)]=E[Y_{111}-(Y_{000}+(Y_{001}-Y_{000})+(Y_{010}-Y_{000}))]=E[(Y_{111}-Y_{010})-(Y_{001}-Y_{000})]=E[(Y_{111}-Y_{001})-(Y_{010}-Y_{000})]=DD$

FUENTE: Elaboración propia.

RECUADRO 5
PANELES DE DATOS

La versión más simple de método basado en panel de datos es la estimación *DD* de diferencias en diferencias, en la que si se reescribe $DD=E[(Y_{111}-Y_{001})-(Y_{010}-Y_{000})]$ como $DD=E[(Y_{111}-Y_{010})-(Y_{001}-Y_{000})]$ se comprueba cómo las componentes idiosincráticas $\alpha_{D=1}$ y $\alpha_{D=0}$ que forman parte respectivamente de $Y_{J1t}$ e $Y_{J0t}$ se cancelan en las diferencias intragrupo que incluye el estimador: $DD=E\{[(Y'_{111}+\alpha_{D=1})-(Y'_{010}+\alpha_{D=1})]-[(Y'_{001}+\alpha_{D=0})-(Y'_{000}+\alpha_{D=0})]\}=E[(Y'_{111}-Y'_{010})-(Y'_{001}-Y'_{000})]$

Un modelo de *efectos fijos* $\alpha_i$ por entidad, del tipo $Y_{it}=\beta_0+X_{it} \cdot \beta_1+\delta \cdot D_{it}+\alpha_i+\varepsilon_{it}$ se puede resolver de dos maneras:

1. incluyendo variables *dummy* por cada una de las entidades –salvo una, para evitar una situación de multicolinearidad perfecta que por la mecánica del modelo ha de ser evitada- de manera que $Y_{it}=\beta_0+X_{it} \cdot \beta_1+\delta \cdot D_{it}+I_{[i=ij]} \cdot \alpha_i+\varepsilon_{it}$
2. ajustando una regresión de variables minoradas por las medias temporales por entidad –lo que se nota con la tilde ~ sobre las variables– siendo entonces que $\tilde{Y}_{it}=\tilde{X}_{it} \cdot \beta_1+\delta \cdot \tilde{D}_{it}+\tilde{\varepsilon}_{it}$

Un modelo que incluya también efectos fijos temporales $\eta_t$ es del tipo $Y_{it}=\beta_0+X_{it} \cdot \beta_1+\delta \cdot D_{it}+\alpha_i+\eta_t+\varepsilon_{it}$

FUENTE: Elaboración propia.

tamiento y control y varias observaciones previas y posteriores a la intervención –lo que daría más credibilidad a la validación que sobre la tendencia común se haga–, en lo que sería una generalización del *DD* utilizando múltiples series temporales organizadas de manera longitudinal en un panel de datos que recoja información de una misma unidad observacional en diferentes momentos (Meyer, 1995).

La gran ventaja de los paneles de datos es que permiten en ciertas circunstancias eliminar el sesgo de una sección cruzada con omisión de variables. En efecto, con una sencilla manipulación algebraica como es la toma de diferencias, los factores de selección no observables de carácter idiosincrático para cada unidad observacional o entidad desaparecen si estos tienen carácter invariante en el tiempo. El arreglo de este método de primeras diferencias puede lograrse también -alternativamente y con la ventaja de poder estimar los efectos de estos inobservables– con el modelo de efectos fijos por entidad, cuya extensión a la dimensión tiempo del panel permite considerar también efectos fijos temporales (recuadro 5).

La lógica tras el supuesto de invariancia es simple: aquello que no cambia no puede ser responsable de impacto alguno. Es por ello que sus violaciones –dada una entidad, variación en el tiempo para los efectos fijos por entidad; y dado un momento, variación entre entidades para los efectos fijos temporales- resultan en la inconsistencia de las estimaciones.

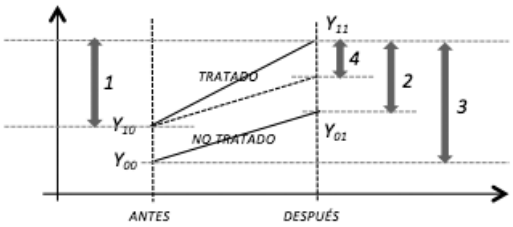
Para clarificar la idea, considérese un panel en el que las distintas entidades fueran lo estados miembros de la UE. Los efectos fijos invariantes por entidad serían aquellos comportamientos idiosincráticos inalterables que hacen que cada uno de los países responda con sus peculiaridades a las diferentes políticas de una manera invariante en el tiempo. Por otra par-

CUADRO 4
ESTIMACIÓN CON ESTRATEGIAS SENCILLAS DE IDENTIFICACIÓN

| Estrategia | Cálculo |
|---|---|
| 1 Pre-post/Antes-después | $Y_{11}$-$Y_{10}$ |
| 2 Sección Cruzada | $Y_{11}$-$Y_{01}$ |
| 3 Comparación de Cohortes (*) | $Y_{11}$-$Y_{00}$ |
| 4 DD/Panel | $(Y_{11}$- $Y_{01})$- $(Y_{10}$-$Y_{00})$ |

donde se ha utilizado la siguiente notación:

| Notación | Antes | Después |
|---|---|---|
| **Tratado** | $Y_{10}$ | $Y_{11}$ |
| **No tratado** | $Y_{00}$ | $Y_{01}$ |




(*) La misma expresión sería a *grosso modo* válida para un experimento social controlado. La gran diferencia es que en la comparación de cohortes no hay aleatorización en la asignación de los agentes a los grupos de tratamiento y control.

FUENTE: Elaboración propia.

te, un efecto fijo temporal invariante entre entidades sería una intervención de política pública comunitaria en un momento dado que afectara por igual a todos y cada uno de los países.

Las evaluaciones sobre el impacto en el largo plazo en variables como educación, salud, criminalidad y salarios de programas de atención infantil reforzada como el *Head Start* de los EE.UU. realizadas por Garces *et al.* (2002) y Deming (2009), son una buena ilustración de la utilidad de los efectos fijos y las amenazas a su validez.

Así, y en este contexto, si bien el nivel educativo de los padres es una variable fácil de controlar, más delicado sería tratar aspectos como el énfasis puesto en la educación. De hecho, las estimaciones de sección cruzada que omiten estos inobservables proporcionan conclusiones tan contraintuitivas como que estos programas son perniciosos en términos de las variables de resultado analizadas.

Una manera inteligente de tratar el problema consiste en explotar en un panel de familias (9) los datos de aquellas con dos hijos en los que uno se matricula en el programa –agente tratado– y el otro no –agente de control-, de manera que el efecto fijo vendría dado para cada una de las familias. La invariancia temporal de los inobservables –o equivalentemente, la tendencia común de los resultados potenciales para los grupos de tratamiento y control– exige que, dada una familia, no se den fenómenos como cambios en el ambiente educativo –que afecten a los resultados de una u otra forma–, influencias –positivas de un niño que atiende al programa sobre su hermano– o favoritismo –hacia uno de los niños–.

Este último factor es la principal amenaza a la validez, al atacar de pleno a la filosofía de emulación de un experimento social controlado que se logra cuando los padres deciden inscribir a un hijo u otro de manera aleatoria, lo que justifica asimismo que no sean útiles para el análisis familias del panel con los dos hijos inscritos o sin inscribir al ser precisamente la variabilidad de los datos la fuente información que se utiliza para la estimación.

Dicha variabilidad puede generarse *a priori* en el momento de diseñar una política mediante diferentes mecanismos: implantación por fases o modulación de la tasa de cobertura (10) –lo que puede darse también de manera no intencionada si hay retrasos accidentales en algunos de los agentes–, administración intermitente, o modulación de la intensidad del programa.

Para finalizar este epígrafe, y a modo de recapitulación antes de comenzar con los venideros de mayor sofisticación en los que se tratarán en detalle las problemáticas de los estudios observacionales –métodos de emparejamiento para selección con observables- y la selección en base a inobservables –variables instrumentales y regresión discontinua-, en el cuadro 4 se sintetiza el cálculo de estimadores con las cuatro estrategias de identificación sencillas descritas hasta el momento, donde en aras a la claridad se ha prescindido del subíndice utilizado en la discusión sobre resultados potenciales dando lugar a la notación para las variables de resultados que se precisa.

## MÉTODOS DE EMPAREJAMIENTO

Sin un proceso de asignación aleatoria que evite el problema de selección y sin experimentos naturales con-

RECUADRO 6
EMPAREJAMIENTO MÚLTIPLE POR GRADO DE PROPENSIÓN

En la estrategia de emparejamiento o *matching* se realiza un ajuste de los datos que persigue que no haya diferencias sistemáticas entre los grupos de tratamiento y control condicionadas a valores de las variables observables, de manera que $E\{[Y_{0i}, D_i=1]-[Y_{0i}, D_i=0]|X_i\}=0$, lo que se conoce como «hipótesis de independencia condicional» y que se suele expresar como $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i|X_i$. Tras aplicar las ponderaciones de ajuste, los grupos de tratamiento y comparación se hallarán balanceados en términos de observables, esto es, las distribuciones de los observables serán las mismas en ambos grupos: $F(X_i|D_i=0)= F(X_i|D_i=1)$.

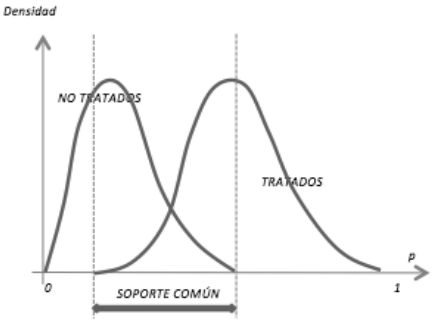
El «Teorema de Rosenbaum y Rubin» (1983), que establece que si $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i|X_i$ entonces $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i|p(x)$ siendo $p(x)=P(D_i=1|X_i=x)$ el grado de propensión, es un importante resultado que legitima la utilidad de este en casos de selección con observables siempre y cuando su estimación –para lo cual en la práctica se usan modelos de elección discreta como el lineal de probabilidad, el *logit* o el *probit*- sea correcta. El hecho de que el grado de propensión sea una variable latente supone una complicación puesto que sus valores no son observables en la realidad y por tanto se desconoce la bondad de su ajuste $\hat{p}$.

La expresión general de un estimador por emparejamiento múltiple es

$$\hat{\delta}=\frac{1}{N_T}\sum_{i\varepsilon T}\left(Y_i-\Sigma_{j\varepsilon C}\, w_{ij}\cdot Y_j\right) \; con \; \Sigma_i w_{ij}=1 \text{ para todo i y } w_{ij}=w(\hat{p}(X_i),\hat{p}(X_j))$$

en donde los pesos $w_{ij}$ miden la comparabilidad de la observación *i*-ésima del grupo de tratamiento *T* con la *j*-ésima de los no tratados *C* y se calculan mediante diferentes algoritmos que para cada agente tratado *i* construyen un comparable "compuesto" considerando los *j* candidatos que se encuentren dentro de una ventana de valores del grado de propensión de manera que, de acuerdo a la métrica de distancia que se defina, se sobreponderen los más próximos y se infraponderen los más lejanos. Los algoritmos de emparejamiento «k-vecinos más cercanos» *Kernel/Epanechnikov*, *Kernel/Gaussiano* o *Kernel/triangular*, por citar algunos), que pueden ser con o sin reemplazamiento –según se pueda o no utilizar varias veces un mismo agente no tratado para construir un contrafactual–, son asintóticamente equivalentes pues en el límite proporcionan emparejamientos exactos.

Es importante matizar que esta estimación se ha de restringir a agentes con observables en el dominio en el que exista «soporte común», que gráficamente es la zona de solape entre los histogramas o funciones de densidad del grado de propensión para los grupos de tratados y no tratados.




FUENTE: Elaboración propia.

vincentes, los «métodos de emparejamiento» (o *matching*) (11) construyen un grupo de comparación que juega el papel del grupo de control de un experimento social buscando a agentes similares a los del grupo de tratamiento entre los no tratados. Para ello, los resultados de estos se ajustan mediante unas ponderaciones para hacerlos así comparables a los de los agentes tratados, tal y como hace el método de sección cruzada –que de hecho es el más elemental de los métodos de emparejamiento– cuando incorpora variables de control por las características observables, pero con la desventaja de estar sometido este a errores de especificación al tratarse de una técnica paramétrica que exige precisar una forma funcional.

En el caso de observable categóricos o discretos, los datos pueden organizarse en forma matricial de manera que el criterio de comparabilidad es inmediato: los observables han de ser idénticos, o lo que es lo mismo, para comparar los resultados de un agente tratado con los de uno no tratado ambos han de estar en una misma celda de la matriz cuyas dimensiones son las variables observables y los valores de los subíndices que identifican sus elementos los rangos o categorías pertinentes.

Este emparejamiento exacto facilita el cálculo sencillo (12) de un «estimador de celda» en el que primero se calculan efectos de tratamiento individual como diferencias de resultados entre agentes tratados y no tratados de una misma celda, y después se ponderan estos impactos individuales con las proporciones de agentes en cada una de ellas de manera que según los pesos que se escojan –proporción

**FIGURA 3**

**COMPROMISO SESGO-PRECISIÓN**

Mayor ancho de banda→más datos utilizados en la estimación→
→mayor poder explicativo/potencia del contraste→
→menor dispersión/varianza/error estándar, mayor precisión/eficiencia→
→mayor sesgo, menor exactitud/consistencia

Menor ancho de banda→menos datos utilizados en la estimación→
→menor poder explicativo/potencia del contraste→
→mayor dispersión/varianza/error estándar, menor precisión/eficiencia→
→menor sesgo, mayor exactitud/consistencia

FUENTE: Elaboración propia.

total, proporción de tratados o proporción de no tratados- se obtendrá el parámetro de tratamiento deseado –*ATE*, *ATET* o *ATEN*, respectivamente–.

En el caso de observables continuos y/o numerosos, el enfoque de celda no es viable puesto que en muchos casos únicamente se dispondrá de agentes o tratados o sin tratar, pero no ambos, resultando vacía la celda, y la solución pasará por hacer un emparejamiento inexacto, en el que se minimice una distancia euclídea entre observables. La alternativa pasa por reducir la dimensionalidad del problema a una variable única o estadístico de síntesis, denominado «grado de propensión» *(propensity score)*, que aglutine la información contenida en los observables y que mida en función de estos la probabilidad de que un agente sea tratado por medio de un modelo de participación.

Un primer método de emparejamiento uno a uno basado en el grado de propensión es el del «vecino más cercano» que toma como contrafactual de un agente tratado aquel agente no tratado con grado de propensión más próximo. Una variante de este método es la que fija un radio de influencia para evitar emparejamientos con contrafactuales muy diferentes.

Por su parte, el «emparejamiento por intervalos/estratos/bloques» es un método sencillo de emparejamiento inexacto en el que se crean varios rangos de valores de grados de propensión, y se pondera por igual a todos los agentes dentro de cada uno de estos rangos, a diferencia de otros métodos más sofisticados que construyen un contrafactual de comparación a partir de varios agentes (emparejamiento múltiple) no tratados similares en términos del grado de propensión y a los que se les da un diferente ponderación en función de su semejanza (recuadro 6, en página anterior).

De lo expuesto se deduce que, en muchos de los métodos de emparejamiento, hay elementos de subjetividad en el criterio de similitud/cercanía/vecindad –¿qué métrica se escoge?; ¿cuántos intervalos/estratos se definen?; ¿cuántos vecinos se toman?; ¿qué tamaño se da a la ventana de emparejamiento? – que se materializan en un diferente «ancho de banda» o «calibre» o «tolerancia», que marca la cantidad de datos utilizada en la estimación, y que ponen de manifiesto una vez más el compromiso entre sesgo –será menor cuanto más parecido sea el contrafactual, lo que se logra tomando menos datos- y precisión –será mayor cuantos más datos se tomen- ubicuo en la inferencia econométrica que se muestra en la figura 3 con un símil de lanzamiento de dardos a una diana.

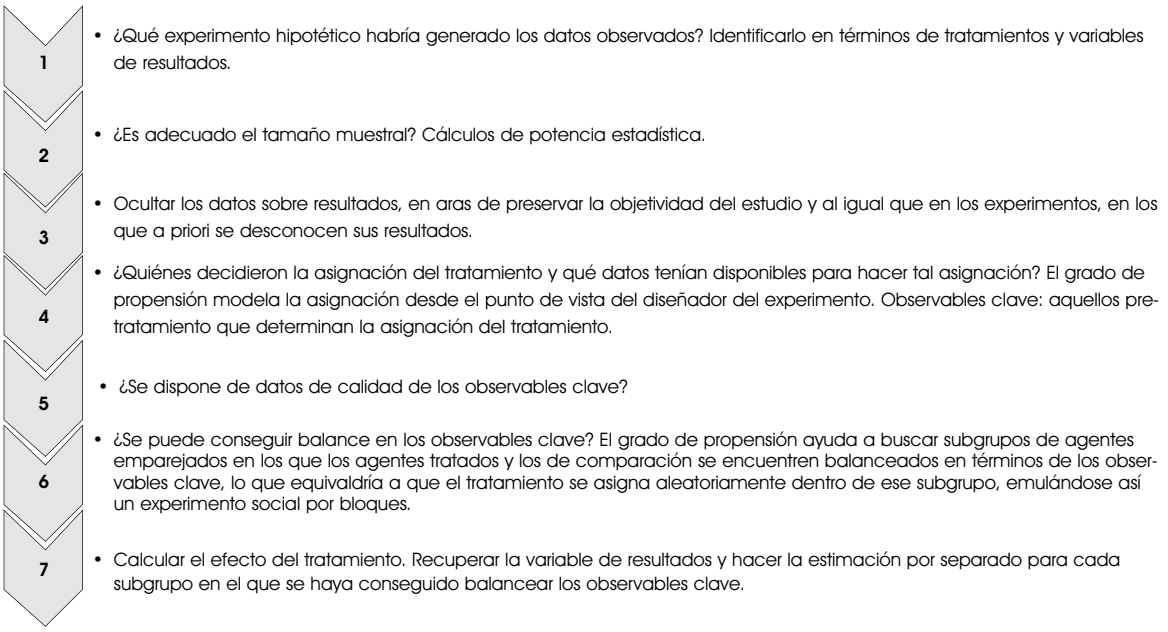
Esta circunstancia puede salvarse con la ponderación inversa de probabilidades (*Inverse Probability Weighting)* que calcula el efecto del tratamiento mediante una regresión ponderada en la que los pesos a aplicar –dependientes del parámetro de tratamiento deseado- a los datos de los agentes tratados o no son función del grado de propensión en un método que es también válido para el caso discreto y que proporciona los mismos resultados que con el cálculo del estimador de celda descrito.

Dos son las hipótesis sobre las que se asienta la validez de la identificación por emparejamiento. La primera es la «hipótesis de independencia condicional»– que al exigir que no haya diferencias sistemáticas entre los agentes tratados y no tratados una vez se condicionan los valores de los observables, garantiza que los contrafactuales escogidos sean adecuados por diferir los agentes emparejados únicamente en el hecho de recibir tratamiento o no. La razón es que controlando a los agentes según características observables en cada subgrupo de valores de estas el tratamiento será independiente de los resultados, lo que equivaldría a una asignación aleatoria.

La segunda es la «hipótesis de soporte común» y exige que, dado un valor de las variables observables, se encuentren agentes tratados y no tratados para su comparación, es decir, haya cierta probabilidad de recibir tratamiento (13). Esto hace inválida la estrategia cuando la selección se hace con criterios determinísticos sobre la base de un valor umbral pa-

**FIGURA 4**

**HOJA DE RUTA PARA ESTUDIOS OBSERVACIONALES**




FUENTE: Elaboración propia a partir de Rubin (2008).

ra cierta variable observable. Para estos casos, resulta de utilidad la estrategia de regresión discontinua que se describe más adelante en este artículo y que permite además solventar el problema de selección en base a inobservables bajo ciertos supuestos.

Si la segunda de las hipótesis es comprobable –pues es una mera condición mecánica que se impone a los datos disponibles–, lamentablemente la primera no lo es, por lo que la selección en base a inobservables que no se acertará a incluir en el modelado del grado de propensión continuará siendo un problema.

Entre las ventajas de los métodos de emparejamiento pueden resaltarse el hecho de que aúnan sofisticación –pues capturan heterogeneidades– y simplificación –pues al ser de tipo no paramétrico no exigen especificación funcional–. Además sólo utilizan una parte de los datos –donde hay soporte común–, por lo que no hay extrapolación alguna de la estimación a regiones donde no hay datos disponibles, lo que es un elemento de refuerzo de la validez interna que evidentemente va en detrimento de la validez externa.

Las desventajas se refieren a las violaciones del supuesto de independencia condicional en casos de selección por inobservables, a las exigencias de disponibilidad de datos –en cuanto a número de agentes y conjunto amplio de observables pre-tratamiento– y a la problemática de la estimación del modelo de participación para el grado de propensión. A tal efecto, los regresores a incluir en este serán aquellos que sugiera la Teoría Económica y que afecten a la participación y a los resultados, lo que excluye en todo caso a variables post-tratamiento o aquellas pre-tratamiento no inmutables –atributos–que se vean afectadas por el mismo.

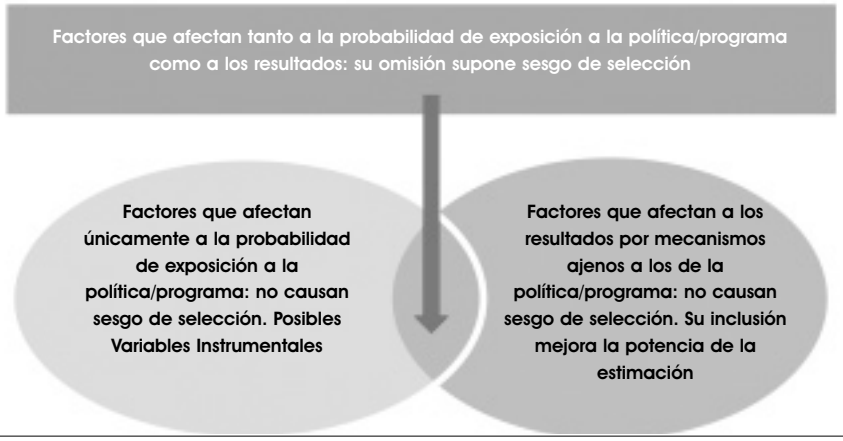
Black y Smith (2004) ilustran algunas de las ventajas de los métodos de emparejamiento al analizar los efectos en los salarios de acudir a una universidad de calidad, medida por un índice que aglutina indicadores parciales de gastos por estudiante, salario de los profesores, nota media de ingreso o tasa de abandonos en el primer año (14). Los problemas que la regresión lineal –que es la técnica de estimación tradicionalmente utilizada en este caso– presenta son la falta de soporte común –pues los buenos alumnos van a las buenas universidades y los menos buenos van a las menos buenas– que dificulta la identificación de contrafactuales, y las complicaciones en la especificación funcional que acaban por violar la hipótesis de linealidad en el condicionamiento a observables en que se basa para la superación de la selección en base a estos. Ello se solventa con un emparejamiento por grado de propensión estimado en base a los observables que marcan la selección de la calidad de la universidad: edad (y su cuadrado), raza, región de nacimiento, medidas de habilidad del estudiante (nota de ingreso y su cuadrado) y características del instituto, los padres, y el entorno familiar.

## Estudios observacionales

Rubin (2008) propone un proceso más o menos secuencial y estructurado (ver figura 4) para llevar a ca-

**FIGURA 5**

**VARIABLES INSTRUMENTALES**




FUENTE: The Magenta Book.

bo estudios observacionales emulando experimentos sociales mediante el artificio de la reconstrucción –con la ayuda del emparejamiento por grado de propensión– del mecanismo de asignación del experimento hipotético que hubiera dado lugar a los datos. En dicho proceso, las etapas que obligarían a abortarlo sin contemplaciones son la segunda -tamaño muestral inadecuado/escaso poder o potencia-, la quinta –falta de datos de calidad sobre los observables clave que determinan si se recibe tratamiento o no– y la sexta -desbalance en observables pre-tratamiento por una incorrecta estimación del grado de propensión–.

Para ilustrar la problemática del desbalance y cómo este es corregido con el emparejamiento por grado de propensión, Rubin utiliza el siguiente ejemplo sencillo. Considérese el caso en el que se quiere evaluar el impacto que las visitas a un doctor por parte de un visitador médico –tratamiento– tienen en el número de prescripciones de un medicamento para la pérdida de peso –variable de resultados–.

En este "experimento", del cual se dispondría de datos suficientes sobre las características de los doctores, las visitas y el número de prescripciones, son los visitadores médicos quienes deciden qué doctor visitar. Al depender sus ingresos de las ventas, lógicamente preferirán a aquellos que tengan consultas de mayor capacidad, un historial de mayor número de prescripciones y que ejerzan las especialidades que más prescriban el tipo de medicamento en cuestión.

Esta preferencia se manifiesta en forma de desbalance o diferencia estadística entre los grupos de doctores visitados o no, lo que amenazaría la consistencia de la estimación. En efecto, si se considera el caso hipotético en que lo que determinase la asignación al grupo de tratamiento –esto es, la probabilidad de visita– fuera únicamente la especialidad del doctor, al tender los visitadores a evitar a los ginecólogos -pues estos no recetan a mujeres embarazadas medicamentos de adelgazamiento-, en el límite esta especialidad no mostraría visita alguna en el juego de datos y, por consiguiente, no se podría estimar el efecto del tratamiento al ser asignado este con una regla determinista –no tratar si el doctor es ginecólogo– en lo que sería una violación de la hipótesis de soporte común.

Sin embargo, para una especialidad como medicina general, al disponer de datos tanto de doctores visitados como no, la asignación del tratamiento replicaría la de un experimento social por lo que se podrían realizar una estimación consistente del impacto de aquel. Esto es precisamente lo que hace el emparejamiento al eliminar el desbalance en observables clave mediante la comparación de doctores visitados y no visitados dentro de una misma banda de valores del grado de propensión estimado en función de los mismos.

## ESTRATEGIAS AVANZADAS CON SELECCIÓN EN BASE A NO OBSERVABLES

En este epígrafe se van a describir dos estrategias de identificación que afrontan el problema de selección en base a no observables, para lo cual, de los métodos expuestos hasta el momento, sólo el DD/panel es válido bajo ciertas hipótesis y con la contrapartida de su exigencia en términos de necesidades de datos.

### Variables instrumentales

La primera de ellas es la de las denominadas «variables instrumentales» o «instrumentos». Dichas variables son aquellas de carácter exógeno –no están directamente relacionadas con el resultado potencial– que son relevantes a la hora de determinar el estado de participación -están relacionadas directamen-

te con la probabilidad de ser tratado-, lo que se puede representar con un diagrama de Venn como el de la figura 5.

El uso de instrumentos ataca directamente el problema de la endogeneidad: si en una estimación de efectos causales en la que la omisión de variables no observables diera lugar a estimaciones inconsistentes se pudiera descomponer la variable de participación en dos partes –una exógena y otra endógena–, la estimación resultante de un modelo en el que como regresor de participación se utilizara únicamente su componente exógena resultaría entonces consistente. La intuición es que se estaría ante un cuasi-experimento donde la variable instrumental, que determina sólo parcialmente el tratamiento, es lo que induce a los agentes a cambiar su estado de participación en aquel. La dificultad estriba en que, por definición, la parte endógena no es observable y, por ende, tampoco la exógena, por lo que en su lugar a lo más que se puede aspirar es a encontrar un proxi que es precisamente la variable instrumental o instrumento y que ha de cumplir las dos condiciones de exogeneidad (o exclusión) y relevancia ya citadas.

A título ilustrativo, un primer ejemplo en el que se usa el trimestre de nacimiento para instrumentar los años de escolarización en el estudio del impacto de leyes de escolarización obligatoria en los salarios (Angrist y Krueger, 2001). Los nacidos en los primeros trimestres alcanzan antes la edad de 16 años de escolarización obligatoria en EE.UU., lo que puede resultar en hasta un año menos de educación. Por tanto, al determinar parcialmente los años de escolarización (relevancia), sin tener nada que ver –excepto vía escolarización– con los salarios (exogeneidad), el trimestre de nacimiento constituye un instrumento adecuado.

Un segundo ejemplo es el uso del sexo de los dos primeros hijos como instrumento válido para medición del impacto en la oferta laboral de las mujeres en los salarios que perciben (Angrist y Evans, 1998): si es igual -dos niños o dos niñas- se tiene tendencia a buscar un tercer bebé, lo que disminuye la oferta laboral (relevancia) sin que este hecho tenga nada que ver con los salarios (exogeneidad).

Se ha de puntualizar que el uso de instrumentos no es novedoso, pues su origen se remonta a la década de los 20 del siglo pasado cuando Philip y Sewal Wright estudiaron la problemática de la determinación de funciones de oferta y demanda a partir de la observación de equilibrios precio-cantidad en los mercados. Para solventar el hecho de que tales equilibrios resultan de una interacción simultánea entre ambas funciones, propusieron el ingenioso enfoque de instrumentar el precio con factores que influyeran en la oferta pero no en la demanda para estimar así esta.

Incidiendo en el carácter cuasi-experimental de la estimación por instrumentos, en condiciones de homogeneidad, y sin problemas de deserción y sustitución, el parámetro de tratamiento que se obtendría sería el *ATE*. En el caso más común de heterogeneidad, se obtendría una aproximación siempre que se dieran dos condiciones adicionales a las de exogeneidad y relevancia como son la de «estabilidad del valor unitario del tratamiento» –condición *SUTVA*– y la de «monotonicidad» que excluye la existencia de agentes que respondan al instrumento de manera «desafiante» o contraria, adictiva o que lo ignoren por completo, en lo que sí que constituye una reinterpretación novedosa de los instrumentos (recuadro 7, en página siguiente).

Por tanto, este aspecto de validez externa pobre es una de las principales desventajas, al tener la estimación un carácter local (*LATE-Local Average Treatment Effect*) no generalizable para toda la población, y válida únicamente para los agentes que marginalmente respondan al instrumento escogido (*compliers* –aquellas familias que optaran por tener un tercer bebé cuando los dos primeros tienen el mismo sexo en el ejemplo de la estimación de la oferta laboral de las mujeres–) que al no ser necesariamente único conlleva que diferentes instrumentos den lugar a diferentes estimaciones. Las otras complicaciones del método se refieren a la dificultad de encontrar instrumentos interesantes desde el punto de vista del análisis –que sean reflejo fiel de un cambio de política– y que cumplan las condiciones adecuadas, más cuando algunas –exogeneidad/exclusión y monotonicidad– no son comprobables.

Finalmente, y de nuevo, aparece el ya citado compromiso entre exactitud/consistencia y precisión/eficiencia que se describió en el caso de los métodos de emparejamiento: con la estimación mediante instrumentos se está sacrificando la precisión/eficiencia en favor de la exactitud/consistencia (menor sesgo), dado que el error estándar del estimador crece al explotarse únicamente la parte de la variabilidad de la variable de interés que está correlacionada con el instrumento.

## Regresión discontinua

La segunda de las estrategias para tratar la selección en base a no observables es la «regresión discontinua» (15), que utiliza una regla administrativa de un programa como es su criterio de elegibilidad para crear así un grupo de comparación en lo que sería un caso de experimento natural con estimación local. Aunque el tratamiento se asigne de manera determinista –«regresión discontinua aguda»– o probabilista –«regresión discontinua borrosa o difusa»– y no aleatoria –como en el RCT– sobre la base de una variable continua («variable de asignación») cuando esta supera un valor umbral (o de corte, o de cualificación o de elegibilidad), si el criterio de elegibilidad está fijado arbitrariamente y no hay posibilidad de comportamientos estratégicos por incentivos en torno al punto de corte («condición de inmanipulabilidad»), entonces los individuos justo por encima y por debajo del mismo serán similares y por tanto las discontinuidades que se observen en los resultados («condición de salto») podrán atribuirse al progra-

RECUADRO 7
ESTIMACIÓN POR VARIABLES INSTRUMENTALES

Los requerimientos formales para un instrumento $Z$ son respectivamente $Cov(Z_i,\varepsilon_i)=0$ (exogeneidad o exclusión) y $Cov(Z_i,D_i)\neq 0$ (relevancia) donde $Z_i$ es la variable instrumental y $D_i$ la de participación para el ajuste de un modelo $Y_i=\beta_0+\delta\cdot D_i+\varepsilon_i$ resultando entonces que el estimador del factor causal en este caso simple es $\hat{\delta}=Cov(Z_i,Y_i)/Cov(Z_i,D_i)$ que se deduce de $Cov(Z_i,Y_i)=Cov(Z_i,\beta_0+\delta\cdot D_i+\varepsilon_i)=\delta\cdot Cov(Z_i,D_i)+Cov(Z_i,\varepsilon_i)$ al tener en cuenta que el segundo sumando es cero por la condición de exogeneidad.

En la práctica, para la estimación mediante instrumentos se utiliza un método bietápico. En una primera etapa se estima un modelo de participación en función de los instrumentos $Z_i$ (tantos como variables con problemas de endogeneidad) y el resto de observables $X_i$ exógenos, de manera que $D_i=\alpha_0+Z_i\cdot\alpha_1+X_i\cdot\alpha_2+u_i$. En una segunda etapa, se estima un segundo modelo de regresión mínimo cuadrático $Y_i=\beta_0+\delta\cdot\hat{D}_i+X_i\cdot\beta+\varepsilon_i$ que utiliza como regresor de participación las estimaciones $\hat{D}_i$ resultantes de la primera etapa y como resto de regresores los observables exógenos $X_i$. Una prueba ácida o de «pulgares arriba» *–rule of thumb–* para validar la relevancia del instrumento es que en la primera etapa el estadístico de Fisher-Snedecor proporcione un valor superior a 10, esto es $F>10$. Complementariamente, $F<10$ sería una indicación de la debilidad (o escasa relevancia) del instrumento, que un problema incluso con tamaños muestrales grandes porque pequeñas desviaciones de la asunción de exogeneidad se exacerbarán por esta debilidad. Para la exogeneidad/exclusión dependiendo del caso podría hacerse alguna comprobación aproximada

En condiciones de heterogeneidad, y cumpliéndose las condiciones de monotonicidad y de estabilidad del valor del tratamiento unitario, además de las de exogeneidad y relevancia, se tiene que la estimación instrumental es una aproximación del tipo de Wald definido en el recuadro 3. En efecto, partiendo del caso simple se tiene que $\hat{\delta}=Cov(Z_i,Y_i)/Cov(Z_i,D_i)=\frac{Cov(Z_i,Y_i)\,/\,Var(Z_i)}{Cov(Z_i,D_i)\,/\,Var(Z_i)}$

de donde el estimador por variables instrumentales resulta ser, $\hat{\delta}=\frac{E(Y_i\mid Z_i=1)-E(Y_i\mid Z_i=0)}{P(D_i=1,Z_i=1)-P(D_i=1,Z_i=0)}$ expresión en la cual la condición de relevancia garantiza que su denominador sea distinto de cero.

La monotonicidad, que tampoco puede comprobarse, se puede expresar formalmente como $P_A+P_N+P_C=1$ o $P_D=0$ donde el sentido de los subíndices *A*, *N*, *C* y *D* se detalla en la siguiente tabla:

| | | $Z_i=0$ | |
|---|---|---|---|
| | | $D_i(Z_i=0)=0$ | $D_i(Z_i=0)=1$ |
| $Z_i=1$ | $D_i(Z_i=0)=0$ | N-Nunca tomadores: deciden no participar en cualquiera de las contingencias | D-Desafiantes: hacen siempre lo contrario |
| | $D_i(Z_i=1)=1$ | C-Cumplidores: hacen lo que dice el programa | A-Adictos (siempre tomadores): participan en el programa salgan o no sorteados |

FUENTE: Elaboración propia.

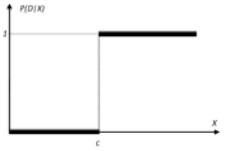
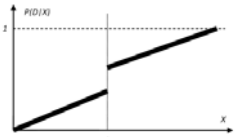
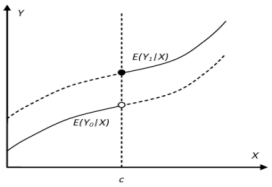
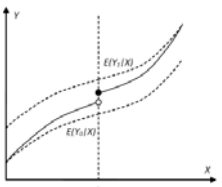
ma en ausencia de otros tratamientos cuyos efectos, de actuar en el mismo umbral, no podrían distinguirse del que se está analizando («condición de suavidad»). Tres son, pues, las condiciones o requisitos de identificación de la estrategia de regresión discontinua: salto, inmanipulabilidad y suavidad.

Ludwig y Miller (2007) utilizan la regresión discontinua para la evaluación del ya citado programa *Head Start* a partir de su elegibilidad en función de la tasa de pobreza, en cuyo valor de corte las diversas variables de resultados del programa (tasa de mortalidad, tasa de finalización de estudios medios, años de escolarización,…) muestran un salto o discontinuidad. La condición de suavidad también se cumple al no haber otros programas sociales que utilicen la misma regla de cualificación, lo que impediría la distinción de los efectos de los distintos programas concurrentes en el umbral.

Finalmente, dado que en el programa no había competencia en el acceso a los fondos por exceder estos a la demanda, la inmanipulabilidad está también garantizada, al igual que la arbitrariedad del corte revelada por la historia del tratamiento recibido, que consistió en capacitar a los 300 condados más pobres –lo que supone un valor umbral del 59,2% para la tasa de pobreza– para cumplimentar las solicitudes al programa. El parámetro que por tanto estima Ludwig es un *ITT* que mide los efectos de hacer el programa disponible, lo que se traduce en mayores fondos y participación que son los mecanismos a través de los cuales se alteran los resultados.

La evaluación del impacto en la mortalidad prematura de tratamientos postnatales administrados a bebés que al nacer no alcanzan un peso mínimo (Almond *et al.*, 2008) es otro ejemplo del uso adecuado de la regresión discontinua al cumplirse sus requisitos de identificación. En primer lugar, la mortalidad es una función decreciente del peso de nacimiento con un salto en el umbral de los 1.500 gramos. Por otra parte, la inmanipulabilidad en el punto de corte está garantizada en tanto en cuanto no tienen sentido comportamientos estratégicos que fuercen el peso de nacimiento

## CUADRO 5
## REGRESIÓN DISCONTINUA

| Tipo de discontinuidad | Aguda (Sharp) | Borrosa o Difusa (Fuzzy) |
|---|---|---|
| **Regla de participación** (c es el valor de corte para la variable de asignación) | Determinista: la probabilidad de recibir tratamiento salta del 0 a 1 en el corte<br>$P(D_i=1 \mid X_i \geq c)-P(D_i=1 \mid X_i<c)=1$<br>  | Probabilista: la probabilidad de recibir tratamiento salta un valor $0<k<1$ en el corte, debido a otros factores generalmente no observados que influyen en la participación aparte de la variable de asignación, como pueden ser errores administrativos<br>$P(D_i=1 \mid X_i \geq c)-P(D_i=1 \mid X_i<c)=k$<br>  |
| **Impacto** | $\delta(c)=\lim_{X_i \downarrow c} E(Y_{1i} \mid X_i=x)-\lim_{X_i \uparrow c} E(Y_{0i} \mid X_i=x)$<br>  | $\delta(c)=\dfrac{\lim_{x_i \downarrow c} E(Y_{1i} \mid \chi_1=x)-\lim_{x_i \uparrow c} E(Y_{0i} \mid X_i=x)}{\lim_{x_i \downarrow c} E(D_i \mid X_i=x)-\lim_{x_i \uparrow c} E(D_i \mid X_i=x)}$<br>  |
| **Estimación no paramétrica** (Estimador de Wald dentro de la ventana de datos $[c-\chi_0, c+\chi_1]$) | $\hat{\delta}=E(Y_{1i} \mid X_i<c+\chi_1)-E(Y_{0i} \mid X_i<c-\chi_0)$ | $\hat{\delta}=\dfrac{E(Y_{1i} \mid X_i<c+\chi_1)-E(Y_{0i} \mid X_i<c-\chi_0)}{E(D_i \mid X_i<c+\chi_1)-E(D_i \mid X_i<c-\chi_0)}$ |
| **Estimación paramétrica** | $Y_i=\beta_0+\beta_1 \cdot (X_i-c)+\delta \cdot D_i+\beta_2 \cdot (X_i-c) \cdot D_i+\varepsilon_i$ | Método bietápico<br>*Etapa 1:* $D_i=\alpha_0+\alpha_1 \cdot I_{[X \geq c]}+u_i$<br>*Etapa 2:* $Y_i=\beta_0+\beta_1 \cdot (X_i-c)+\delta \cdot \hat{D}_i+\varepsilon_i$ |

FUENTE: Elaboración propia, gráficos: García Núnez (2011), Imbens y Lemieux (2008)

al valor de cualificación. Finalmente, el que otras variables potencialmente influyentes (edad de la madre, nivel socioeconómico, cuidados prenatales,…) muestren suavidad (ausencia de saltos) en torno al valor de corte dan credibilidad a que el salto en la variable de resultados se deba al tratamiento.

La estimación por regresión discontinua puede hacerse con dos enfoques, uno no paramétrico y otro paramétrico, cuya formalización para los dos diseños posibles –agudo y borroso o difuso- se muestra en el cuadro 5. En el primero, los resultados a ambos lados del umbral se promedian dentro de un calibre o ancho de banda de manera que cuanto mayor sea este, mayor número de datos se usan en la estimación que será menos dispersa pero menos exacta en una nueva manifestación del compromiso consistencia-eficiencia. En el segundo, se realiza el ajuste de todos los datos de resultados disponibles con una forma funcional dada, lo que siendo *a priori* atractivo, puede dar lugar a que el ajuste en torno al umbral -que es donde se está haciendo la estimación localmente- sea inexacto.

Las principales ventajas de la regresión discontinua es que esta proporciona una estimación consistente de tipo *LATE* en la discontinuidad cuando se aplica correctamente y que, en su calidad de cuasi-experimento que no requiere por tanto de aleatorización previa, está exenta de la preocupación sobre aspectos éticos.

En cuanto a las desventajas, además de su escasa validez externa, se refieren a la exogeneidad de la variable de asignación (incluso cuando el umbral de corte parece arbitrario, todavía puede darse selección en base a no observables); la plausibilidad de las condiciones de suavidad (que es parcialmente comprobable analizando en torno al punto de corte los covariantes previos al programa) e inmanipulabilidad (comprobable analizando la función de densidad de la variable de asignación alrededor del corte); el compromiso implícito en los métodos no paramétricos; y la problemática de la especificación funcional en los paramétricos.

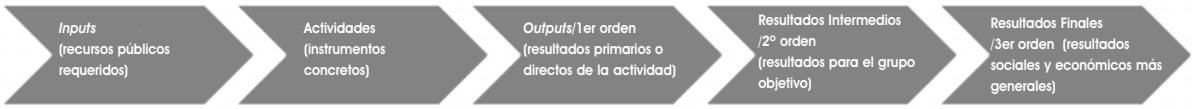
## REFLEXIONES FINALES: ¿OTRA BOUTADE DE LA TORRE DE MARFIL?

Una vez descritas las diferentes técnicas microeconométricas utilizadas en la evaluación de políticas públicas conviene hacer algunas consideraciones finales sobre las mismas:

**La Evaluación de Programas es una disciplina eminentemente cuantitativa**, hasta el punto que hace

**FIGURA 6**

**MODELO LÓGICO DE CAUSALIDAD**




FUENTE: The Magenta Book

suyo el dicho «lo que no se puede medir, no se puede evaluar»(16).

Un punto de partida es la selección de variables de resultados relevantes a distintos niveles, para lo cual son de utilidad los modelos lógicos de secuenciación de la Kellogg Foundation (ver figura 6) en función de cuya complejidad se diagnostica a qué nivel es posible la evaluación cuantitativa del impacto.

**Del carácter cuantitativo se deriva la necesidad de invertir en datos** sobre las variables que se quieran medir para que estos estén disponibles en tiempo y forma.

Las fuentes de datos pueden ser encuestas diseñadas *ad-hoc* o registros administrativos ya existentes. Suele ser habitual combinar/triangular fuentes múltiples, lo que si técnicamente es en principio sencillo, afronta barreras más complejas de tipo organizativo -que se manifiestan en las reticencias a compartir datos de quien se siente propietario de ellos (17)– o ético –consideraciones sobre protección– que deberán ser abordadas.

Desde el punto de vista técnico, es importante la manera en que se organizan los datos (sección cruzada, sección cruzada repetida, panel,…), la cantidad de los mismos (que afecta a aspectos de poder explicativo/potencia del contraste y significación estadística y económica de los estudios que se realicen), su calidad y las amenazas que suponen su explotación indiscriminada (minería de datos o «empirismo ciego» según García Nuñez, 2011).

**No existe técnica perfecta o ideal para el procesado de los datos en busca de relaciones de causalidad** (estrategias de identificación), puesto que cada una tiene su alcance (análisis *ex -ante* vs *ex-post*), hipótesis de partida y tipo de información que proporciona (*ITT*, *LATE*,…) que la hacen válida para problemáticas concretas y que justifican un análisis caso por caso.

Más en concreto, y en un artículo seminal en el que Lalonde (1986) analiza diferentes evaluaciones de un mismo programa comparándolas con el resultado de un experimento controlado, las diferencias con este sirven para llamar la atención sobre peligros que como los errores de especificación amenazan a las diversas técnicas y ponen en valor el *RCT*.

**Las técnicas descritas admiten mejoras** para su aplicación en la asignación de tratamientos socialmente óptimos –lo que exige un posicionamiento normativo- en función del agente tratado -mediante el uso del *targeting* o el *profiling*– y para la captura de fenómenos como la selección dinámica –que tuviera en cuenta los efectos acumulativos que se dan al recibir secuencialmente los tratamientos de diferentes programas–, la heterogeneidad de tratamientos o la incorporación de efectos de equilibrio general (Durán, 2004).

Como consecuencia de los puntos anteriores, ha de quedar claro que la identificación de relaciones causales en el contexto de una filosofía de políticas basadas en evidencias es una tarea ardua que a menudo forzará a la toma decisiones con información incompleta debido a limitaciones -como la imposibilidad de contrastar las hipótesis de las diferentes estrategias o los conflictos entre validez interna y externa- que habrán de ser tenidas en cuenta en todo momento, pero que de por sí no restan valor al ejercicio de evaluación.

## Las tres dimensiones de la evaluación: cuantitativa, económica y de procesos

Pero la evaluación de políticas públicas va más allá del impacto empírico (que la política sea eficaz en términos de resultados) cubierto por la Evaluación de Programas, teniendo otras dos dimensiones con las que esta se interrelaciona estrechamente: la económica (que la política sea eficiente en base a criterios más o menos cuantitativos como el coste-beneficio o el coste-efectividad) y la de procesos (que la política se ejecute tal y como se concibió, en un análisis de carácter eminentemente descriptivo).

La relación entre la evaluación de impacto y la económica se manifiesta en que, por un lado, la primera (que se realiza *ex-post*) aporta información que se incorpora en la metodología de la segunda (que se puede realizar *ex-ante* o *ex-post*), y por otro, la segunda proporciona el marco conceptual que, utilizando conceptos de Economía del Bienestar como el excedente del consumidor o del productor, sirve de guía para formular las preguntas adecuadas y escoger las variables de resultados relevantes para la primera.

Con esta triple dimensión empírica, económica y de procesos a la vista, en la evaluación de políticas pú-

blicas en sentido amplio se han de tener en cuenta las siguientes consideraciones:

**Como fase dentro de un ciclo de vida que se retroalimenta, ha de tener en cuenta la totalidad del proceso** (Moreno-Torres, 2012). De ahí la importancia de concebir *ex-ante* y durante la fase de diseño la estrategia de evaluación para una política (18).

Así, y en el caso de evaluaciones de impacto, la incorporación de la evaluación en el proceso de diseño de un programa o política puede hacerse controlando su asignación en una fase piloto de alcance parcial (para la realización de un *RCT*, la introducción por fases u operación intermitente por oleadas para el uso de métodos *DD*/panel o mediante reglas objetivas de asignación con criterios de elegibilidad que faciliten la evaluación por regresión discontinua) o estableciendo diferentes niveles de exposición en caso de que el alcance del programa sea total. Aún y con todo, se han presentado estrategias (pre-post, experimento natural, variables instrumentales y emparejamiento) que permitirían en principio realizar *a posteriori* una evaluación no planificada *a priori*.

**Como actividad costosa que consume recursos múltiples, ha de estar sometida al criterio coste-beneficio** de manera que se lleve a cabo cuando sus beneficios –derivados del uso de sus conclusiones y recomendaciones- excedan a sus costes teniendo en cuenta criterios de proporcionalidad con el riesgo, escala y perfil del programa en cuestión.

La gestión de esos recursos tiene una dimensión temporal, derivada de hechos como que los efectos de las políticas pueden tardar en manifestarse o que pueda ser recomendable que una evaluación de impacto venga precedida de una de proceso.

**Como práctica, ha de tener una aproximación mixta cuantitativo-cualitativa**, pese a la preferencia general de la objetividad de lo cuantitativo frente a la subjetividad de lo cualitativo. El asesoramiento cualitativo de participantes, operadores o expertos resulta de la mayor utilidad, aun reconociendo su propensión a sesgos como el exceso de optimismo o el «postureo».

Feinstein (2007) plantea un enfoque pragmático orientado a resultados y la rendición de cuentas, en el que distingue las políticas concebidas como tratamientos (programas o proyectos) de las que diseñan un marco normativo no reducible a una simple suma de aquellos, para las cuales prescribe una triangulación de métodos. Asimismo resalta el papel de la institucionalización (19) como remedio para solventar el problema de escasez de incentivos que a menudo dificulta que del proceso evaluatorio se extraiga un aprendizaje.

## Evaluación de políticas industriales

Si existe cierta tradición de uso de los métodos microeconométricos de Evaluación de Programas en el ámbito de las políticas sociales, aunque en grado menor, también empieza a acumularse y compartirse (20) experiencia y conocimiento en la evaluación de políticas industriales. A título ilustrativo, y por citar algunos ejemplos:

**Reindustrialización**. Criscuolo *et al.* (2012) evaluan el impacto en variables como empleo, inversion, productividad y número de plantas de un programa de asistencia regional selectiva que subvenciona inversiones industriales en áreas desfavorecidas del Reino Unido. Para ello instrumentan la participación en el programa de una planta en un momento dado con el nivel de subvención a la inversión máxima disponible en el área en dicho momento, del que pueden beneficiarse sólo las plantas localizadas en ciertas áreas (relevancia) y que viene determinado por las normas de la Unión Europea sobre ayudas de Estado (exogeneidad) que son actualizadas periódicamente en lo que viene a constituir un experimento natural.

**Desarrollo y dinamización de *clusters***. Schmiedeberg (2010) hace una revisión metodológica de las diferentes técnicas utilizadas en la que además de modelos econométricos, contempla otras herramientas como la evaluación de procesos, el estudio de casos, los enfoques sistémicos (modelos *input-output*, análisis de redes y *benchmarking*) y la evaluación económica coste-beneficio.

**Ayudas a la I+D**. Aerts *et al.* (2006) analizan el estado del arte de su evaluación econométrica focalizando en aspectos como el desplazamiento de la inversión privada por la pública, la adicionalidad de los resultados en innovación o la adicionalidad «comportamental» –cambios de comportamiento inducidos– en la investigación colaborativa.

**Innovación y emprendimiento.** CPB (2012) es un informe que analiza las estrategias para la evaluación econométrica de diferentes instrumentos para su fomento y apoyo utilizados en los Países Bajos. Así, se propone el uso de métodos de emparejamiento para un programa de créditos a la innovación de alto riesgo en productos, servicios y proyectos en el que los proyectos se seleccionan en un procedimiento que incluye una cualificación objetiva –lo que permite modelar el grado de propensión– y la opinión de un comité, dando lugar a una base de proyectos ″dudosos″ a partir de los cuales construir contrafactuales. Para un programa de deducciones fiscales a la innovación que exige un mínimo de horas anuales de actividad innovadora se propone un diseño de regresión discontinua para la estimación de impactos entorno al umbral exigido. Un modelo de panel con efectos fijos por entidad y temporales se sugiere para un programa de apoyo a proyectos en partenariado entre centros públicos de I+D y empresas innovadoras de sectores estratégicos, en una primera aproximación con las debilidades propias de tomar como contrafactual compañías del mismo sector no participantes en el partenariado o compañías del partenariado no participantes en el proyecto.

**Eficiencia energética y control de emisiones.** Martin *et al.* (2011) utilizan un método de panel para eva-

luar el impacto de la tasa por cambio climático en el consumo energético de plantas manufactureras del Reino Unido, comparando las sujetas a la totalidad de la misma con aquellas beneficiarias de deducciones por haber adquirido un compromiso voluntario de ahorro en energía o emisiones, en lo que da lugar a situaciones de autoselección que los autores solventan con una variable instrumental basada en la elegibilidad para la suscripción de aquel, que es razonablemente exógena por basarse en el cumplimiento de la regulación ambiental previa a la exacción de la tasa.

Por su parte, Allcott (2011) analiza el impacto en términos de ahorro de electricidad del envío a consumidores de un informe personalizado en el que se les comunica su situación relativa en términos de consumo, se les categoriza y se les recomienda una serie de medidas de ahorro, basándose en el experimento llevado a cabo por una compañía eléctrica de EE.UU. en el que aleatoriamente se decidió enviar o no tal informe a los clientes de su base. Asimismo utiliza un análisis de regresión discontinua para evaluar el impacto de las categorizaciones.

Se trata en todos estos casos de instrumentos o programas que suponen el uso de palancas concretas que activan resortes de causalidad relativamente sencillos y total o parcialmente exentos de consecuencias inesperadas o efectos indirectos de orden superior. Desafortunadamente, en el ámbito de lo industrial estos son bastante habituales y se presentan en forma de desplazamiento/*crowding out* –un resultado positivo es contrarrestado por uno negativo–, sustitución –el efecto sobre un agente es a costa de otros agentes–, *spillover/leakage* –el programa beneficia a agentes no pertenecientes al grupo destinatario–, *deadweight* –los resultados de la política se hubieran obtenido igualmente sin intervención- o aglomeración/*crowding*.

A ello se une la creciente complejidad de lo que podríamos denominar «Nueva Política Industrial», más integrada y sistémica que la tradicional enfocada a sectores y –en lo que en términos de motivación supone una mayor sensibilidad hacia los fallos de gobierno que hacia los fallos de mercado– con un menor nivel de intervención que se orienta al desarrollo del marco institucional y de clima de negocios adecuado, que facilite la cooperación entre agentes (para la superación de asimetrías de información y de fallos de coordinación) y la alineación de incentivos entre los sectores público y privado (mediante fórmulas de partenariado).

De ahí que estén cobrando fuerza conceptos como «Metaevaluación» (evaluación de evaluaciones para el caso de intervenciones complejas, multiobjetivo y a gran escala, usando por ejemplo modelos de simulación) *y Policy Learning/Intelligence* que constituyen un nuevo enfoque en el que la evaluación de políticas, de ser un mero ejercicio de auditoría *a posteriori*, pasa a convertirse en una fuerza transformadora (Aranguren *et al.*, 2013), lo que exige desarrollar una cultura de la evaluación que involucre a los agentes, incorpore sistemas de monitorización, seguimiento y difusión, adopte enfoques mixtos cuantitativo-cualitativos y asegure la independencia del proceso evitando situaciones juez-parte que tan perversas resultan.

Una de las técnicas para este aprendizaje transformador y la introducción experimental a pequeña escala de nuevas políticas sobre la base del contraste de hipótesis sobre diferentes alternativas (*Policy Experimentation*) es el experimento social o *RCT* ampliamente discutido en este artículo. Sirva como ejemplo la experiencia que en el area de Manchester lideró en 2009 el organismo público de fomento de la innovación en el Reino Unido (21) en relación con un esquema de créditos para la promoción de partenariados innovadores entre PYMEs y proveedores de servicios creativos. Un experimento en el que los créditos se asignaron conforme a los resultados de una lotería permitió comprobar la adicionalidad de la política (recibir el crédito aumentó la probabilidad de que una PYME se asociara con un proveedor de servicios creativos), para validar así un modelo lógico de causalidad sobre el que legitimar un escalado del programa. Frente a un esquema tradicional en que los apoyos públicos se conceden una vez hechas evaluaciones *ex-ante* de los proyectos a subvencionar, la aleatorización del tratamiento demostró también la ventaja de evitar los costes de aquel en términos de consumo de recursos y problemas de información asimétrica.

Si bien el uso de experimentos de políticas públicas está menos difundido en Europa en comparación con EE.UU. –la aleatorización en la asignación de los tratamientos es una rareza, por administrarse precisamente a los más necesitados–, en el caso de las políticas industriales, y con todas las salvaguardas ya apuntadas, habría argumentos en favor de su uso con mayor intensidad que en el caso de políticas sociales, tales como sus menores implicaciones éticas –al ser las unidades experimentales empresas en lugar de individuos o familias– y la mayor sensibilidad desde el punto de vista de las finanzas públicas –al ser por lo general mayores los desembolsos por beneficiario–.

## Un comentario final

Las técnicas descritas, lejos de ser un ejercicio estéril de lucimiento académico, son herramientas que, desde el reconocimiento de sus limitaciones y complicaciones, están listas y a disposición de los *policy makers* para ser aplicadas en la construcción del nuevo paradigma que en la actualidad demandan las políticas públicas, uno de cuyos aspectos clave ha de ser el desarrollo de una cultura de la evaluación para lo cual este artículo ha pretendido ser una modesta contribución.

**(*) El autor agradece al programa Fulbright y al Ministerio de Industria, Energía y Turismo (MINETUR) el patrocinio de sus estudios en la Universidad de Chicago durante el curso 2011-2012 que han posibilitado la redacción de este artículo.**

## NOTAS

(1) El heurístico e intuitivo *Homo Psicologicus* frente al racional *Homo Economicus* de los libros de texto.

(2) El de Blundel y Costa (2009) es un excelente artículo de similar alcance que este pero con mayor detalle y formalismo, al igual que el de Imbens y Wooldridge (2008). Otro buen artículo de carácter introductorio publicado en español es el de García Núñez (2011). Tratados de mayor calado, algunos de ellos eminentemente prácticos, son los libros que se enumeran en listado de bibliografía recomendada incluida en este artículo.

(3) Heckman (2010) explora una tercera vía que sirva de puente entre los métodos estructurales y los de resultados potenciales.

(4) Un ejemplo de estimación mediante serie temporal interrumpida podría ser la evaluación del impacto de una medida que busque aumentar la transparencia en el mercado minorista de productos petrolíferos, como es el caso del Geoportal de Hidrocarburos del MINETUR que está disponible en su web y que ofrece información en tiempo real de los precios de todas las estaciones de servicio de España –véase la nota incluida en el número 386 de Economía Industrial–. Como contrafactual se tomaría la tendencia de precios ajustada con un «Modelo de Corrección de Errores» (MCE) habitualmente utilizado en la literatura sobre series temporales de precios de productos petrolíferos y que captura realidades como el fenómeno conocido de «cohetes y plumas» según el cual hay una asimetría en las dinámicas de precios de los productos derivados del petróleo, que se ajustan casi simultáneamente a las subidas del precio del crudo pero que en contra lo hacen con retraso a las bajadas.

(5) ¿Pasar por la Universidad de Chicago garantiza el Premio Nobel o más bien es que es más probable que de la Universidad de Chicago salga un Premio Nobel precisamente porque a ella acuden los más preclaros talentos atraídos por su prestigio y filtrados por el proceso de admisión? El autor aprovecha para aclarar que, si bien sí que se vio atraído por el prestigio de dicha Universidad, ni mucho menos se considera un preclaro talento y, desafortunadamente, tampoco candidato al Premio Nobel.

(6) Para ello Card y Krueger utilizan dos series temporales de datos de empleo en restaurantes de comida rápida ubicados en zonas fronterizas de dos jurisdicciones –estados de New Jersey y Pennsylvania- que plausiblemente presentan una tendencia común, y en una en las cuales –NJ- en un momento dado se aumenta el salario mínimo. Curiosamente se llega a una conclusión que resulta contraintuitiva desde el punto de vista de las teorías neoclásicas sobre el mercado del trabajo que pronostican que aumentos del salario mínimo se traducen en descensos en el empleo.

(7) Grogger analiza datos de criminalidad en zonas "limpias" a las que se prohíbe a las pandillas acceder en comparación con otras no sometidas a dicha prohibición y suficientemente alejadas de aquellas para evitar influencias o "contagios" de aquellas, por lo que en principio constituirían un grupo de control razonable.

(8) Las estimaciones del impacto en los salarios de la capacitación de desempleados resultan sesgadas al alza cuando los administradores del programa lo ofrecen preferentemente a aquellos que antes del mismo están en condiciones especialmente desfavorecidas en términos de salarios.

(9) El modelo, en el que el subíndice *j* que identifica a cada hijo haría el papel de dimensión temporal, sería el siguiente: $Y_{ij} = \beta_0 + \delta \cdot HS_{ij} + \beta_1 \cdot PRE_{ij} + X_{ij} \cdot \beta_2 + \alpha_i + \varepsilon_{ij}$ donde $HS_{ij}$ es la variable binaria de participación en *Head Start*, $PRE_{ij}$ la de participación en otros programas de preescolar, $X_{ij}$ es un vector de variables de control ($\beta_2$ tendría también carácter vectorial) *y* $\gamma_i$ es el efecto fijo para la familia *i*.

(10) En las que se basan las evaluaciones del programa sanitario Seguro Popular de México -uno de los que recientemente más interés ha atraído desde el punto de vista de la evaluación- y que complementan las realizadas a través de experimentos cuya validez externa es débil por haberse realizado sólo en áreas rurales con un elevado grado de cobertura sanitario no representativo de la realidad media del país.

(11) Todd (2006) es un excelente trabajo sobre los métodos de emparejamiento.

(12) $\hat{\delta} = \Sigma_{j=1}^{N} w_j \cdot \hat{\delta}_j$ donde *N* es el número de celdas y si $n_j$ es el número de observaciones en la celda j-ésima, $n_C$ en el grupo de control y $n_T$ el de tratamiento, y $n = n_C + n_T$, como pesos $w_j$ se tomarían $w_j = n_j/n$ para el *ATE*, $w_j = n_j/n_T$ para el *ATET y* $w_j = n_j/n_C$ para el *ATEN*.

(13) De manera que $0 < P(D_i = 1 | X_i = x) \equiv p(x) < 1$ para todo *x* observado.

(14) En lo que sería una manifestación de lo que se conoce como teorema de Tiebout de "voto con los pies".

(15) Imbens y Lemieux (2007), discuten aspectos teórico-prácticos sobre regresión discontinua.

(16) «If you can't measure it, you can't evaluate it» (Profesor Lalonde).

(17) Las iniciativas de «Reutilización de Información de Sector Público» *(RISP),* en un reconocimiento de que los administrados son los últimos propietarios de los datos en manos de las Administraciones, buscan poner estos a disposición de aquellos para la creación de valor a través del desarrollo de aplicaciones y servicios basados en los mismos en el contexto de un nuevo paradigma conocido como «Datos Abiertos»/*Open Data* u *Open Government*.

(18) Tal y como recogen las metodologías de análisis de impacto de la Comisión Europea.

(19) En España la Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (AEVAL) tiene por misión «La promoción y realización de evaluaciones y análisis de impacto de las políticas y programas públicos, así como el impulso de la gestión de las calidad de los servicios, favoreciendo el uso racional de los recursos y la rendición de cuentas a la ciudadanía».

(20) En foros de intercambios de buenas prácticas como por ejemplo el Grupo de Expertos en Evaluación de Políticas Industriales del Comité de Industria, Innovación y Emprendimiento (CIIE) de la Dirección de Ciencia, Tecnología e Industria de la OCDE.

(21) National Endowment for Science, Technology and the Arts (NESTA).

## BIBLIOGRAFÍA

AERTS, K., CZARNITZKI, D. y FIER, A. (2006): «Econometric evaluation of public R&D policies: current state of the art». *Working Paper*, K.U. Leuven.

ALLCOTT, H. (2011): «Social norms and energy conservation». *Journal of Public Economics*.

ALMOND, D., DOYLE, J. J. JR., KOWALSKI, A. E. y WILLIAMS, H. (2008): «Estimating Marginal Returns To Medical Care: Evidence From At-Risk Newborns». *NBER Working Paper Series*.

ANGRIST, J. D. y EVANS, W. N. (1998): «Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size». *The American Economic Review*.

ANGRIST, J. D. y KRUEGER, A. B. (2001): «Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments». *Journal of Economic Perspectives*.

ARANGUREN, M. J., MAGRO, E. y WILSON, J. R. (2013): «La evaluación como herramienta para transformar las políticas de competitividad». *Economía Industrial*, nº 387.

ASHENFELTER, O. (1978): «Estimating the Effect of Training Programs on Earnings». *The Review of Economics and Statistics.*

BLACK, D. A. y SMITH, J. A. (2004): «How robust is the evidence on the effects of college quality? Evidence from matching». *Journal of Econometrics*.

BLUNDELL, R. y COSTA DIAS, M. (2009): «Alternative Approaches to Evaluation in Empirical Microeconomics». *The Journal of Human Resources*.

CARD, D. y KRUEGER, A. B. (1994): «Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania». *The American Economic Review*.

CPB NETHERLANDS BUREAU FOR ECONOMIC POLICY ANALISYS, (2012). «Dare to measure: Evaluation designs for industrial policy in The Netherlands» Final report of the Impact Evaluation Expert Working Group.

CRISCUOLO, C., MARTIN, R., OVERMAN, H. y VAN REENEN, J. (2012): «The Causal Effects of an Industrial Policy». *NBER Working Paper* Series.

DEMING, D. (2009): «Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start». *American Economic Journal: Applied Economics*

DURÁN, C. (2004): «Evaluación microeconométrica de las políticas públicas de empleo: aspectos metodológicos». *Hacienda Pública Española/Revista de Economía Pública*. Instituto de Estudios Fiscales.

FEINSTEIN, O. (2007): «Evaluación pragmática de Políticas Públicas». *ICE, Revista de Economía*,nº 836 monográfico sobre Evaluación de Políticas Públicas, mayo-junio 2007.

GARCES, E., THOMAS, D. y CURRIE, J. (2002): «Longer-Term Effects of Head Start». *American Economic Review*.

GARCÍA NÚÑEZ, L. (2011): «Econometría de evaluación de impacto». *PUCP. Economía*, vol. XXXIV, nº 67, semestre enero-junio 2011

GROGGER, J. (2002): «The Effects of Civil Gang Injunctions on Reported Violent Crime: Evidence from Los Angeles County». *The Journal of Law and Economics*.

HECKMAN, J. J. (1976): «The Common Estructure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models». *Annals of Economic and Social Measurement.* NBER.

HECKMAN, J. J. (1979) «Sample Selection Bias as a Specification Error». *Econometrica*.

HECKMAN, J. J. (2000): «Datos Microeconómicos, Heterogeneidad y la Evaluación de Políticas Públicas». Fundación Nobel, 2003. Revista Asturiana de Economía.

HECKMAN, J. J. (2010): «Building bridges between structural and program evaluation approaches to evaluating policy». *NBER Working Paper Series*.

HECKMAN, J. J. y SMITH, J. A. (1995): «Assesing the Case for Social Experiments». *Journal of Economic Perspectives*.

HOLLAND, P. W. y RUBIN, D. B. (1988): «Causal Inference in Retrospective Studies» *Evaluation Review.*

IMBENS, G. M. y LEMIEUX, T. (2007): «Regression Discontinuity Designs: A Guide to Practice». *Journal of Econometrics.*

IMBENS, G. M. y WOOLDRIDGE, J. M. (2008): «Recent Developments in the Econometrics of Program Evaluation». *NBER Working Paper Series.*

LALONDE, R. J. (1986): «Evaluating the Econometric Evaluations of Training Programs with Experimental Data». *The American Economic Review*.

LUDWIG, J. y MILLER, D. L. (2007): «Does Head Start Improve Children's Life Chances?: Evidence from a Regression Discontinuity Design». *The Quarterly Journal of Economics*.

MARTIN, R., DE PREUX, L. B. y WAGNER, U. J. (2011): «The impacts of the Climate Change Levy on manufacturing: evidence from microdata». Centre for Economic Performance. The London School of Economics and Political Science.

MEYER, B. D. (1995): «Natural and Quasi Experiments in Economics». *Journal of Business and Economic Statistics.*

MORENO-TORRES GÁLVEZ, A. (2012): «Un Marco Conceptual para el Análisis de Políticas Públicas». *Economía Industrial,* nº 385.

ROY, A. D. (1951): «Some Thoughts on the Distribution of Earnings». *Oxford Economic Papers.*

RUBIN, D. B. (2008): «For Objective Causal Inference, Design Trumps Analysis». *The Annals of Applied Statistics.*

SCHMIEDEBERG, C. (2010): «Evaluation of Cluster Policy: A Methodological Overview». *Evaluation*.

TODD, P. E: (2006). «Matching Estimators».

## Bibliografía recomendada

«*Mostly Harmless Econometrics. An Empiricist´s Companion*». JOSHUA D. ANGRIST y JÖRN-STEFFEN PISCHKE, 2009. Princeton University Press.

«*Microeconometrics. Methods and Applications*». A. COLIN CAMERON y PRAVIN K. TRIVEDI, 2005. Cambridge University Press.

«*Causality. Models, Reasoning and Inference*». JUDEA PERL, 2000. Cambridge University Press.

«*The Magenta Book. Guidance for Evaluation*». 2011. HM Treasury.

«*The Green Book. Appraisal and Evaluation in Central Government*». 2011, 2003. HM Treasury.

«*Handbook on Impact Evaluation. Quantitative Methods and Practices*». SHAHIDUR R. KHANDKER, GAYATRI B. KOOLWAL Y HUSSAIN A. SAMAD, 2010. The World Bank.

«*Program Evaluation Methods: Measurement and attribution of Program results*». Third Edition, 1998. Treasury Board of Canada.