

Los experimentos aleatorios y la nueva economía del desarrollo*

MOHAMED SOUAL
Universidad Grenoble Alpes

Resumen

El artículo analiza los experimentos aleatorios controlados (RCT)** que sus promotores quieren que sean una microeconomía del desarrollo basada en la producción de evidencia (Evidence Based Approach) mediante el uso de evaluaciones ex-ante de programas de desarrollo. El auge de estos RTC coincide con la reconfiguración del campo de la economía del desarrollo y el lugar que ocupa el análisis empírico. Apoyado en una amplia base mediática, el método ocupa un lugar preponderante en el mercado de los peritajes. Comparando grupos a los que se aplican los tratamientos con los grupos control, los RCT buscan inferir, sin bases teóricas, causalidades “indiscutibles” que plantean problemas epistemológicos y no desprovistos de sesgos estadísticos. Además de los problemas de validez externa que el método sugiere, sus aporías estadísticas y lagunas epistemológicas nublan la metodología, lo que contribuye a deconstruir lo que ha sido el *gold standart* que exigen sus promotores.

Palabras clave: experimentos controlados, microeconomía del desarrollo, causalidad, confractualidad, efectos

Clasificación JEL: O12, O17, C93, C21

Random controlled experiments and the new development economics

Abstract

The article analyzes Random Controlled Trial (RTC) experiments that their promoters wish to be seen as a development microeconomy based on the production of evidence (Evidence based approach) through the use of evaluations ex ante of development programmes. The rise of these RTC coincides with the reconfiguration of the field of development economics and the role played by empirical analysis. Supported by a wide mediatic base, the method occupies a major place in the expert market. By comparing groups which have received these treatments to control groups, the RTC seek to infer without theoretical basis, -undeniable- causalities which pose epistemological problems which at the same time show statistical features. In addition to problems of external validity occasioned by the method, its statistical deficiency and epistemological weaknesses cloud the methodology and together deconstruct the *gold standard* which its promoters demand.

Keywords: Experiments controled, development microeconomics, causality, confractuality, effects

JEL: O12, O17, C93, C21

* La traducción del francés al castellano estuvo a cargo de ENERGEIA

** Expérimentations aléatoires contrôlées (EAC), Random Controlled Trial (RCT).

Los experimentos aleatorios y la nueva economía de desarrollo

Dentro del marco de la reconfiguración de su campo y sus contenidos, iniciado en los años ochenta bajo el impulso del FMI y el Banco Mundial, la economía del desarrollo cada vez más ha reservado un lugar para el análisis empírico, ganando de esta manera de los aportes de la econometría tanto de la evaluación de las políticas económicas y la compatibilidad con el crecimiento, como a la economía de la educación, la salud y/o el trabajo.

El enfoque de las evaluaciones aleatorias de las políticas de desarrollo promovidas por Abhijit Banerjee y Esther Duflo, es protagonista en este momento de inflexión y que para muchos sugiere abonar sobre la revolución de la credibilidad (Angrist y Pischke, 2010) que emerge de un nuevo desarrollo económico. En oposición a las técnicas de regresión en la macroeconomía, este enfoque, que recurre a las experimentaciones controladas y que consiste en hacer tests sobre una gran cantidad de programas vigentes sobre poblaciones beneficiadas comparadas con otras poblaciones y que son fuente testimonio para identificar los efectos causantes. A este respecto, se aboga por una refundación micro-económica de políticas públicas que tiene por objeto comprender el desarrollo dentro de una línea de ingeniería que trata de la lucha contra la pobreza.

Definida de esta forma, se afirma que sus contribuciones son originales a partir de un modelo que reúne los ingredientes de los métodos de Holland, Neymar y Rubin y que estudian las cuestiones de la causalidad y contrafactualidad (parte 1). Al examinar esta modelación, se pone en rango de *gold standard*, y donde se aprecia que contiene los límites que realmente preocupan al déficit teórico, con sesgos estadísticos que suponía habían sido eliminados precisamente por los métodos aplicados, así como a los efectos ambiguos de procedencia contrafactual. Examinando cada uno de estos proyectos, la inferencia causal propuesta por el método experimental refleja los aportes de una primacía indebida concedida a lo empírico. (Parte 2).

De tal forma, esta investigación propone además de describir la metodología empleada por los autores, examinar algunos de los aportes de la visión micro individual con el objetivo de enriquecer la micro y la mesoeconomía. Asimismo, el presente artículo se ha propuesto mostrar los límites, las aporías y las fallas epistémicas, silenciadas frecuentemente, y ha contribuido en la deconstrucción de la pretensión forzada de la validez interna conferida por la eliminación de los medios de selección mostrando subsecuentemente que los problemas de generalización que conocen los experimentos aleatorios controlados (RCT) son, precisamente, debidos, por un lado a la debilidad de la validez interna, y a la ausencia de teoría, por el otro.

Principios de la causalidad estadística y dependencia contrafactual

“La buena noticia [...] es que es posible mejorar la gobernanza y las políticas sin cambiar la estructura social y política existente” (Duflo y Banerjee 2012, p. 271). Considerando las dimensiones políticas e institucionales, microsociales como datos, el proyecto del Jameel Poverty Action Laboratory (J-PAL) del MIT se posiciona deliberadamente en una óptica que privilegia la racionalidad individual. Por tal motivo, se basa en la extensión del campo de la economía del desarrollo del método de los ensayos clínicos, se usa en la medicina y busca identificar los efectos de un programa frente a cadenas de causalidad.

Este método busca formular preguntas básicas en los campos de la educación, de la salud, de microcrédito o de la actividad de producción agrícola.

Porque los participantes en un programa no son comparables a los no participantes, y no se dispone de información que permita saber lo que les habría ocurrido si ellos no hubieran participado en el programa, es complicado separar el efecto causal del programa de otros factores (Banerjee, Duflo 2009, p.693). Por consiguiente, el objetivo es comparar la situación de los beneficiarios de un programa a la situación que ellos habrían vivido si ellos no hubieran estado expuestos a ella. La dificultad consiste, por lo tanto, en constituir un grupo de control adecuado cuando no se puede a la vez aplicar un programa y no aplicarlo a la misma persona y al mismo tiempo.

Para los promotores de este método, las evaluaciones realizadas a partir de datos de observación no experimentales tienen sesgos estadísticos. También, sólo la evaluación por experimentación aleatorizada permite liberarse de esos sesgos de selección proponiendo asignar el programa de forma aleatoria mediante sorteo al grupo que se beneficiará del programa tras ser evaluado, o a un grupo control. Uno constituye el grupo de tratamiento y el otro grupo de control, dicho también contrafactual. “Desde luego es imposible, afirman Banerjee y Duflo, identificar los efectos del tratamiento de forma individual para cada persona.

Pero una experiencia aleatorizada, haciendo variar un solo factor a la vez (el programa), nos permite obtener una estimación válida del efecto promedio de un tratamiento para una población dada” (2009:693).

El recurso a este procedimiento permite – según el J-PAL – salir del sueño dogmático –las ideologías– para entrar en la era del pragmatismo y de la cientificidad. Además, se opone contra la economía política de la ayuda al desarrollo, dando la espalda a las dos grandes teorías de la pobreza global. Aquella defendida por Jeffrey Sachs emite la idea de una “trampa de la pobreza” en la que los países pobres estarían encerrados y de la cual no podrían salir más que con la ayuda de envíos de fondos masivos. La segunda, impulsada por William Easterly, rechaza la idea de cualquier tipo de ayuda, responsable según él de corrupción generalizada, de infantilización de los pobres y de debilitamiento de la gobernanza local. Según los promotores de experiencias aleatorias controladas, en lugar de examinar la ayuda internacional, programa por programa, a fin de precisar cuáles son los programas que funcionan y los que no, los expertos se enfrentan a “grandes preguntas” como: ¿Cuál es la causa última de la pobreza? ¿Qué confianza se puede conceder al mercado? ¿La democracia es una cosa buena para los pobres? ¿La ayuda internacional tiene un papel que desempeñar?

Jeffrey Sachs (2005) estima que ni el mercado ni la democracia podrán sacar a los países pobres de su trampa, la ayuda internacional es la clave en la medida que es capaz de iniciar un círculo virtuoso a través de la inversión que aumente la productividad y los ingresos y creen las condiciones para un crecimiento sostenido.

Por otro lado, William Easterly considera que la ayuda a los países del sur no puede resolver el problema del subdesarrollo. En dos obras consagradas al tema (Easterly, 2006, 2009), el autor sostiene que cuando el mercado genera incentivos adaptados y estimula el espíritu de iniciativa mientras que la ayuda propaga la corrupción y los comportamientos rentistas, los agentes consiguen encontrar soluciones para resolver sus problemas.

A diferencia de Craig Burnside y David Dollar (1997) quienes consideran que existe un fuerte vínculo entre crecimiento y la ayuda en los países donde las instituciones son sólidas, él subraya que los países que reciben más ayuda no registran un progreso más rápido de su PIB.

Contra esta perspectiva macroeconómica, Duflo y su equipo sostienen que la evaluación aleatoria del campo se impone como una alternativa para evaluar el impacto de las acciones públicas. Para sustentar esto, Duflo no duda en afirmar que “No es posible, aislar los mecanismos profundos del crecimiento económico como única guía de crecimiento de una centena de países. Toda variable puede ser causa o efecto, o bien podría ser explicada por una tercera variable correlacionada a las otras dos. [...] Hay que someterse a la evidencia. Incluso con dos millones de regresiones, no lograremos descubrir el secreto del crecimiento a partir de una base de datos sobre la experiencia pasada de un centenar de países.” (Duflo 2009, pp. 23-24)

En este contexto, los vínculos entre los niveles macro y micro plantean dificultades debido a la frecuencia reducida de los datos y su cuestionable fiabilidad en los países pobres. Con muestras pequeñas, costos reducidos y preguntas hechas muy específicas, el método experimental ofrece una colección de datos originales, y es lo que permite gastar más dinero por variable y por observación contrario a lo que se hace con grandes encuestas polivalentes de hogares y canastas de consumo o de las empresas que deben contabilizar las variaciones debidas a los cambios de política o inducidas por el mercado (Duflo, Banerjee 2009, p.700). La microeconomía es según ellos, el único campo apropiado en el tratamiento de la pobreza y para la promoción de mejoras al margen de medios de evaluación rigurosos y hasta producir efectos acumulativos, los pobres sufren de sesgo de racionalidad y de incoherencia temporal. El énfasis debe situarse sobre el comportamiento de los pobres y sobre la manera con la cual toman sus decisiones (Favereau, 2016).

Esta nueva configuración de la economía del desarrollo ha vuelto a poner de moda las cuestiones de método e interroga sobre la relación entre la teoría y la experiencia. Así, defender las pruebas «irrefutables», resaltar las inferencias causales “indiscutibles” por un trabajo empírico recolectando, almacenando y haciendo que hablen los datos por sí mismo, constituye, lo que algunos no dudarían en llamar “el giro empírico” de la economía. Mientras que era inimaginable publicar un artículo utilizando métodos econométricos empíricos que no apoyan un modelo teórico, la nueva economía del desarrollo se libera de esta exigencia metodológica. Más que nada aspira a las posibilidades de probar las teorías por la experimentación y el control de condiciones de interacción de fenómenos con el objetivo de despejar las causas o los determinantes de los efectos.

Esta transformación fundamental de la relación entre la teoría y lo empírico plantea la cuestión del control del contenido de las teorías por lo empírico y el delicado paso de la descripción estadística a la generalización inductiva.

Los procesos de elaboración del **conocimiento** **saberes** y los procedimientos de su validación son una preocupación epistemológica permanente. Los verdaderos grandes problemas de la economía del desarrollo, que son también los grandes problemas de la economía en su conjunto, se derivan de la adhesión al individualismo metodológico en un marco de elección “racional” de análisis económico induciendo la elaboración de políticas que con frecuencia no son ni únicamente teorías ni exclusivamente empíricas. Solamente resta constatar que la economía del desarrollo, más integrada hoy a la economía convencional durante los últimos treinta años, ha entrado en una fase fuertemente empírica, habiendo coincidido con una mejor y más accesible disponibilidad de datos y una mayor potencia de cálculo.

En este contexto, la nueva economía del desarrollo ha dado lugar a un vivo debate o han sido convocadas cuestiones relativas al estatus de las proposiciones observacionales con respecto de la teoría. El simposio reunido por el *Economic and Political Weekly* (2005),

alrededor de cuatro más prestigiados economistas del desarrollo de la India: Dilip Mookherjee, Kaushik Basu, Abhijit Banerjee y Pranab Bardhan, aporta a este importante debate metodológico varios elementos que se revisan. Mookherjee, caracteriza cuatro etapas de la investigación:

La primera etapa, corresponde a la descripción empírica, la segunda etapa a la formulación de la teoría, la tercera etapa a los ensayos y la cuarta etapa a la previsión y a la evaluación de las políticas. A lo largo de los últimos quince años, se ha hecho hincapié en la tercera etapa. Actualmente, la economía del desarrollo está principalmente basada en el ensayo, en particular en la dificultad de deducir la causalidad de la correlación. Este es el tema de los debates clásicos, como aquél entre Keynes y Tinbergen¹, destacado por Mookherjee y Bernard Walliser.

Lo anterior es particularmente cierto para dos áreas de la literatura empírica: aquella que se concentra en la búsqueda de variables instrumentales y la que utiliza un modelo experimental de evaluación aleatoria. Llevan al límite la metodología empírica en economía donde Banerjee explica, en parte, por qué la economía del desarrollo es respetada en la disciplina de la ciencia económica: “Creo que el hecho de que no haya defensa particular para un trabajo empírico que utilice datos de países en desarrollo es la razón por la cual los más brillantes y los mejores entre los estudiantes graduados en economía llegan ahora al campo” (*Economic and Political Weekly*, 2005, pages 4340-4344). Según Mookherjee y Bardhan, esta atención a los datos empíricos podría dirigirse a una serie de fenómenos teóricos interesantes, tales como “las implicaciones dinámicas de las imperfecciones del mercado de factores para los agentes individuales, la interacción de los procesos económicos con las instituciones políticas y las normas sociales” (Mookherjee, op.cit.), o bien “... cómo ir más allá de los modelos de oligopolio de equilibrio parcial existentes para generar una teoría viable de los precios de los factores y de la distribución de los ingresos en una economía generalmente oligopolística, la naturaleza de las dinámicas de transición (por oposición a la estadística comparativa) a medida que nos alejamos del equilibrio tradicional de bajo nivel” (Bardhan, op.cit).

Además, esta es una acusación especialmente fundada contra las concepciones experimentales aleatorias, el hecho mismo de que la inferencia sea “pulcra”, a saber, un parámetro específico de control, vuelve problemática cualquier generalización a partir de esta inferencia. Esto es lo que dice particularmente Kaushik Basu, en su contribución a ese simposio.

Pero en la defensa de este método empírico, Banerjee establece la elección entre un elemento de prueba donde la inferencia es clara, en comparación con muchas otras, de lugares y partes diferentes, donde la inferencia es problemática y donde no existe ninguna duda: “Aunque muchas regresiones de baja calidad dicen lo mismo, no hay ninguna razón para que las pruebas de alta calidad se vuelvan inútiles. Después de todo, la misma fuente de sesgo podría afectar a todos los resultados de baja calidad”, aceptando sin embargo que muchos experimentos aleatorios realizados en diferentes lugares y en diferentes momentos permiten reforzar la confianza en sus resultados.

Ninguno de los protagonistas dirá que el uso de una simple correlación para deducir una causalidad debería ser la forma de operar de la economía del desarrollo empírico. Por otra parte, un acuerdo parece surgir sobre el hecho de que el entorno, incluso suficientemente

¹ Keynes critique Tinbergen et son ouvrage *Statistical Testing of Business-Cycle Theories* de postuler par le calcul économétrique l'état futur de l'économie peut être calculé à partir de données statistiques passées.

controlado y generando los datos para permitir una inferencia neta, no es válido más que para ese ambiente y no otro. El papel de la teoría para ayudar a sacar el máximo provecho de este acuerdo tampoco es refutado.

Críticas y controversias a propósito del *Gold Standard*

Esther Duflo y Abhijit Banerjee (2012, p.38) consideran que, debido a la ignorancia, a la ideología y a la inercia, lo que ellos nombran las tres “i”, muchas políticas de desarrollo han fracasado. La evaluación eficaz de los programas de desarrollo debe pasar por la experimentación aleatoria controlada consagrada como el *Gold Standard* metodológico que permite producir resultados fiables y sólidos aportando pruebas inequívocas para la eliminación de todos los sesgos de selección.

El *Gold Standard* atribuye a los RCT (*Randomise Controller Trial*) el incremento que han tenido las evaluaciones de impacto durante las dos últimas décadas, en particular en cuanto al impacto de las políticas de desarrollo. Igualmente provienen de la reivindicación de la superioridad intrínseca de los RCT sobre cualquier otro método (Duflo *et al.*, 2007, citado por Bédécarrats *et al* 2013).

Este atributo también se deriva de la importancia dada a los artículos académicos utilizando los RCT en las revistas más prestigiosas y el atractivo se pronuncia cada vez más sobre los doctorandos. La atribución a Esther Duflo de la medalla Clark que distingue a los economistas de menos de cuarenta años que ejercen en los Estados Unidos, la cátedra “saberes contra la pobreza” en el Colegio de Francia y la audiencia por el Congreso Americano de Dean Karla, presidente y fundador de Innovations for Pauvert Action (IPA), una ONG de recaudación de fondos para los RCT (Bédécarrats *et al.*, 2013, p.118), han contribuido para forjar ese prestigio.

El éxito rotundo de los RCT se debe de igual manera a sus funciones tanto retórica como performativa. Su función retórica se encarna en toda la energía desplegada en todos los foros, incluido principalmente el foro político, con la convicción de que “conocer, es comunicar” (Walliser *et al*, 1988). Comunicar a la comunidad académica para la difusión de métodos y herramientas necesarias (redes asociadas, revistas consagradas, formación e investigación, cátedra, etc.). Difundir ante prescriptores o financiadores (ONG de desarrollo, organismos de desarrollo y financiamiento o grandes fundaciones proveedoras de fondos) y en los medios de comunicación para la divulgación de lo que Walliser llama los *Folk Economiques* (obras destinadas al público en general, artículos de prensa, sitios especializados o foros de alto impacto mediático). Lo que ha hecho escribir a Vincent Guillin (2013): “cualquiera que haya sido testimonio de la difusión en economía del desarrollo en el paisaje intelectual francés y el entusiasmo que ha suscitado en las instancias encargadas de las políticas públicas no dejará de haber sido marcado por la capa de superficie mediática tomada por este movimiento, y por su capacidad de imponer a un público una agenda de investigación muy estructurada.” Su función performativa deriva naturalmente de la constatación de Michel Callon: “La ciencia económica en el sentido de su desempeño, da forma, más de lo que explica el funcionamiento.” Las revistas, las ONG asociadas y los organismos que financian los RCT se impregnan de ellos antes de difundirlos. Mantiene igualmente “lo profundo de la tecnocratización de la representación estadística económica y social, asumida por los animadores del J-PAL quienes reivindican el estatus de “co-experimentadores” (Bardet *et al.*, 2012, p.186).

El Límite de los Randomized Controlled Trial

El *gold standard* atribuye a los RCT de apoyarse en malentendidos: la aleatorización provee una prueba equitativa igualando todo excepto el tratamiento y permite entonces una estimación precisa del tratamiento solo; lo cual es necesario para resolver la selección sin sesgo que aísla la inferencia causal estudiada; y finalmente esta inferencia estadística sólo requiere la comparación de dos medias (diferencia de las medias despejadas para los grupos de tratamiento y de control). Ninguna de estas declaraciones es cierta. Ahora bien, sin “remontar los fenómenos a sus principios y descubrir las estructuras que les regulan” (Frank, 2007), la falta de vínculo entre los RCT y otros conocimientos científicos dificulta su uso fuera del contexto exacto en el que son llevados a cabo (Deaton, Cartwright, 2018).

Los datos que están fuera están lejos de ser ideales para estimar los parámetros estructurales de los modelos de comportamiento. Esto vuelve difícil la generalización de los resultados de una experiencia a otra o la utilización de experiencias para identificar los parámetros estructurales invariantes de una política necesaria para la evaluación de la política (Heckman, 2010, citado por Roux, 2015, p.15).

Los experimentos aleatorios controlados entran en lo que se puede llamar “las estrategias de actores y de construcción de nociones de prueba y de variedad” que analizamos bajo la lectura de Michel Callon y Bruno Latour en “la traducción y el actor red” o bien según Bruno Latour en “la ciencia en acción” o la fórmula actor- red que designa al mismo tiempo una red heterogénea de intereses alineados entre sí y el procedimiento que lleva eventualmente a la producción de un artefacto sociotécnico. Este marco teórico se basa en ciertas nociones clave. Una es justamente la distinción entre el concepto de -actor- central, del que dependen otros elementos de los cuales la voluntad se traduce en su propio lenguaje, y el de -actuante- (**actant**), que designa a su vez a los humanos y los no humanos de una misma red. Otra noción clave es la -controversia-, que es una condición necesaria para la constitución de red y para su traducción por el actor (Mahil et Tremblay, 2015, pp.234-237). En resumen, un contexto o la producción científica se asemeja a un bricolaje doble del arte consumido de la negociación.

En este contexto de revolución neoliberal donde la ley cede el lugar al programa, prosperando sobre las bases de un nuevo ideal normativo y buscando la realización eficaz de objetivos mensurables en lugar de la obediencia a leyes justas, los constructos sociales sostenidos por la revolución digital y los modelados, son el nuevo imaginario pretendiendo ser herramientas del gobierno y de la construcción de la prueba (Alain Supiot, 2015, pp.47-48).

Herramientas de “prueba y de gobierno” según la fórmula de Alain Desrosières (2008) las experimentaciones aleatorias son portadoras a la vez de promesas de científicidad y de renovación de la acción social (Labrousse, 2010); sus limitaciones prácticas puestas en la historia de la experimentación social y del gobierno son señaladas tanto a nivel teórico, estadístico, de validez externa y de generalización de resultados como sobre el plan epistemológico relativo a la inferencia causal en economía inducida por la contrafacticidad, a la cual volveremos más adelante.

Las controversias sobre los límites de estos métodos encuentran su origen en lo que Deaton (2009) y Rodrik (2008) han resaltado como la ausencia de un modelo teórico subyacente. Para Deaton, los RCT acumulan hechos y no conocimiento, y el problema no es tanto lo que funciona sino cómo lo hace. Rodrik no dice otra cosa cuando escribe: “se experimenta, pero ¿qué se aprende?”. Los límites de la validez interna debido a la

metodología de la aleatorización que permite eliminar los sesgos de selección son ampliamente destacados por la literatura dedicada a ellos (Favereau, 2015; Labrousse, 2010; Jatteau, 2016; l'Horty et Petit, 2011; Zamora, 2011). Encuentran, según las hipótesis que formulamos, su origen en la ausencia de ese marco teórico. Sigue siendo la aleatorización misma todavía provista de aporías de orden estadístico señaladas también ampliamente (Heckman y Smith, 1995; Manski, 1997; Fougère, 2012). Estas aporías estadísticas generan al menos tres órdenes de sesgos que la aleatorización supone eliminar (Fougère, 2012).

Escasez de efectivos. Por razones de costes o de tiempo, los investigadores experimentales recurren a la reducción de efectivos detonando en resultados que pueden ser imprecisos ya que la desviación en estándar puede ser demasiado amplia para llegar a una conclusión suficientemente fiable.

Efecto Hawthorne. El efecto Hawthorn, es bien conocido por los psicólogos laborales y describe la situación en la que los resultados de un experimento no se deben a los factores experimentales, sino al hecho de que los sujetos son conscientes de participar en una experiencia en la que ellos están a prueba, propiciando una mayor motivación, y en todo caso, el comportamiento de los sujetos estudiados se modifica por el hecho de haber sido o no elegidos para participar en el experimento y no por el dispositivo o el mismo programa. Esto se acerca al efecto Pigmalión, que se observa con los alumnos cuyos resultados mejoran por el simple hecho de que el profesor se interesa más en ellos.

Los efectos de eliminación natural y sustitución. Estos sesgos son frecuentes, en el caso de experiencias que se desarrollan en una escala de tiempo largo, en el hecho de que los miembros del grupo de tratamiento o del grupo de control deciden abandonar la experiencia (efecto de eliminación natural). Es el caso, por ejemplo, de la maleta de los padres, experimento llevado a cabo por el fondo de experimentación de la juventud, manejada en el marco de las misiones confiadas a Martin Hirsch, Alto Comisionado para la Juventud. El efecto de sustitución concierne generalmente a los miembros de los grupos de control que se someten a otros programas, como aquellos destinados al regreso del empleo. A ello se añaden los problemas de permeabilidad o, según Devaux-Spatarakis, los individuos del grupo control pueden ponerse de acuerdo para tener acceso al tratamiento, o a la inversa, los individuos del grupo de prueba escapar de él (Devaux-Spatarakis, 2014).

Además, las influencias políticas y las interacciones investigadoras experimentales y los socios pueden plantear problemas si recurren a los -bricolajes- que afectan el crédito adjudicado a los RCT. Es el caso de la experimentación en Camboya relativa a la evaluación de un programa de micro seguros (Quentin et Guerin, 2013) o de la experiencia tan elogiada y citada sobre los vermífugos, o los asociados locales no autorizando la aleatorización en la asignación de escuelas y han preferido una asignación por orden alfabético (Labrousse, 2016).

La experimentación ha servido igualmente como transposición de un método científico fuera de su marco de origen, donde la noción de experimentación introducida en la revisión de la Constitución en 2003 fue utilizada cuando el dispositivo de evaluación estaba en curso: una ley que generaliza el Ingreso de Solidaridad Activa (RSA), y en efecto, ha sido promulgada sin esperar los resultados de esta experimentación (Massieu, 2001). Este fue el caso también con el programa Progreso de ayuda enfocado en México (Faulkner, 2012).

La inferencia causal en economía: contra facticidad y prueba controlada aleatoria

La interdicción formulada por David Hume de deducir juicios sobre lo que debe ser a partir de proposiciones fácticas sobre lo que es, ha servido con frecuencia en la filosofía de las ciencias para caracterizar el trabajo científico, particularmente en lo que concierne a las aproximaciones normativas.

El Conocer y el Cómo son cuestiones comúnmente reactualizadas en el conjunto de los subcampos de la economía (econometría, economía experimental, teoría de la decisión, modelización macroeconómica, etc.) (Courgeau, 1997). El problema en la nueva economía del desarrollo se basa en las experimentaciones aleatorias incrementadas por la falta de datos, en un momento donde un individuo no puede ser a la vez beneficiario y no beneficiario de una política, volviendo inobservable el impacto causal de esta última. Este problema es conocido bajo el nombre de –situaciones inobservables derivadas del análisis contrafactual–.

El acercamiento contrafactual en la materia de causalidad en las ciencias sociales teorizado por David Lewis (1973) utiliza una teoría de mundos posibles que considera que son tan concretos y reales como este en el cual vivimos, y es lo que permite hablar de eventos contrafactuales: aunque no se hayan producido, son tan concretos como los que observamos.

Así, para señalar la desproporción de causas y efectos, Pascal había escrito en los *Pensamientos*: “si la nariz de Cleopatra hubiera sido más corta, toda la faz de la Tierra habría cambiado.” Es en el caso figurado donde, por ejemplo, si A se considera como el antecedente de un enunciado y C su consecuente, el enunciado contrafactual toma la forma siguiente:

« $A > C$ » o si “A hubiera sido el caso, entonces C habría sido el caso, donde”; $>$ es el símbolo del condicional contrafactual. Todo el razonamiento se va a basar en la condición de verdad lewisiana siguiente: Se considerará que si un evento C causa otro evento E y si se pregunta si existe un mundo posible en el cual E es el caso, aunque C no lo sea y si ese mundo es más similar al mundo actual que no es el mundo posible, lo más cercano del mundo actual en el que ni C ni E son el caso.

Aplicación de la contra facticidad a las experiencias aleatorias controladas

Este enfoque parte de una hipótesis que desde el principio no es evaluable. En efecto, si se aplica a un individuo u un tratamiento para evaluar, si su respuesta a este es $Y_t(u)$, o si se le aplica un tratamiento clásico c llamado control, siendo $Y_c(u)$ la respuesta, se mide el efecto del tratamiento de control por la diferencia: $Y_t(u) - Y_c(u)$. Como se indicó más arriba, no se puede aplicar simultáneamente al mismo individuo los dos tratamientos, y si se les aplica enseguida uno al otro, cuando es posible, generalmente ya no se puede hablar de causalidad. Sin embargo, si se constituyen dos grupos al azar, así como dos individuos tomados en cada uno, u_1 y u_2 , siendo estrictamente comparables, a excepción del tratamiento dado, el efecto causal del tratamiento sería: $Y_t(u_1) - Y_c(u_2)$.

Para pasar del individuo al grupo, el modelo de Ruben es convocado: El modelo causal contrafactual toma sus raíces en los trabajos de Jerzy Neyman y Fisher citados por Holland (1986, pp. 945-960) en el marco de experimentos aleatorios; fue ampliado posteriormente por Rubin (citado por Holland) proponiendo un marco más general aplicable a las ciencias sociales y que Holland llama «modelo causal de Rubin». Este modelo causal permite medir lo que se llama «los resultados potenciales en función de la exposición del agente al tratamiento».

Estos dos pilares son, por un lado, el parámetro de interés, que representa el fenómeno económico que busca analizar; y por otro, el análisis causal que sigue al impacto impuesto al parámetro de interés a través de la modificación de los factores de influencia. Si el sujeto ha sido sometido al tratamiento, se tiene $T_i = 1$, y si no ha recibido el tratamiento, se califica $T_i = 0$, de modo que cada i perteneciente al grupo de tratamiento o al grupo de control, se observa un Y_{0i} o Y_{1i} , pero no los dos. Como no se puede conocer la distribución de los efectos del tratamiento, $Y_{1i} - Y_{0i}$, gracias a la aleatorización se podrá observar la diferencia entre el resultado medio en los tratamientos y el resultado medio en los controles, que se anota:

$$E(Y_i | T_i = 1) - E(Y_i | T_i = 0) \quad (1)$$

Esta diferencia puede ser dividida en dos términos:

$$E(Y_i | T_i = 1) - E(Y_i | T_i = 0) = [E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1)] - [E(Y_{0i} | T_i = 1) - E(Y_{0i} | T_i = 0)] \quad (2)$$

Sea:

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) = E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1) \quad (3)$$

De manera que la diferencia entre el resultado medio tratado y el resultado medio no tratado representa la diferencia entre los dos resultados observables en el grupo de tratamiento y en el grupo de control, el último término a la derecha no es observable.

Entonces, consideramos la media de la diferencia, en lugar de la diferencia de medias que actualmente se encuentra a la derecha de (3) que es una estimación del efecto medio del tratamiento en las personas tratadas, la cual, según el tratamiento y los controles solo difiere en la aleatorización, es una estimación del efecto medio del tratamiento para todos:

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) = E(Y_{1i} - Y_{0i} | T_i = 1) \quad (4)$$

Así, los problemas basados en la generalización de los resultados de los experimentos aleatorios controlados provienen del hecho de que su objetivo no es comprender la estructura subyacente del sistema de relaciones que genera esos resultados, sino únicamente medir el impacto estadístico para probar el efecto causal atribuible a la intervención. En la ausencia de teoría que puede imponer correcciones, el poder causal que puede justificar la inferencia se debilita ya que los RCT aportan información sobre la media de los efectos del tratamiento, no identifican otras características de la distribución.

En uno de los textos más recientes dedicados a la causalidad, que tomamos de Bernard Paulre (2007), que se pronuncia más que nada sobre el propósito de la noción de causa, B. Saint-Sernin (2002), citado por Paulre quien observa: “Cuanto más se reflexiona sobre la noción de causa, más parece oscura. Sin embargo, su función es clara: sirve para reconstituir los encadenamientos, para inscribir los acontecimientos a primera vista separados en una trama temporal.” Según esta idea, es la sucesión temporal de los eventos la que constituye, según B. Saint-Sernin, la manifestación típica de una explicación causal: “de Platón y Aristóteles hasta hoy, el concepto de causalidad ocupa un lugar central en la búsqueda de la

verdad, porque tiene por función volver inteligibles los encadenamientos temporales.” Siempre según Paulre y esta vez citando a J. Piaget, es la idea de producción, utilizada en un sentido muy general, es decir, que no se limita necesariamente a la idea física de causa eficiente, la cual es privilegiada si se asocia a la idea de necesidad: “La causalidad aparece en todos sus niveles y bajo todas sus formas como implicando a la vez la producción de una novedad, porque el efecto es nuevo con relación a la causa y que hay por lo tanto transformación, y por otra parte, una relación necesaria sin la cual no se pueda hablar de causalidad.”

Así dos posiciones parecen desprenderse: ya sea que se sitúe en los objetos físicos, y la causalidad sería entonces una característica ontológica del vínculo que podría tenerse entre dos objetos (¿cuál es la naturaleza de una causa?), o bien, se posiciona al nivel epistemológico de las representaciones del mundo físico, se busca explicar y no revelar principios intrínsecos (¿cómo es posible inferir la existencia de una causa a partir de la observación?).

Las dos características de la economía experimental residen en dos aspectos primordiales: el -control-, que se refiere a la capacidad del experimentador de dominar lo que ocurre durante la experiencia para producir datos de buena calidad, y la -replicación-, que permite producir un experimento análogo a fin de probar la robustez de los resultados. Estas dos características garantizan en general la producción de conocimientos experimentales. Pero esto es precisamente lo que remite a los problemas de validez interna y externa planteados de forma recurrente por las críticas a los RCT.

Jatteau (2013) evoca la presión a la publicación que condiciona la carrera de los investigadores para que un resultado positivo tenga más posibilidades de ser publicado, así como el problema del conflicto de intereses en la medida o la distinción evaluador/experimentador en las pruebas aleatorias no es tomado en cuenta sobre el plano científico, con la esperanza de descubrir de forma objetiva el efecto de un programa. Todas esas manipulaciones añadidas a los sesgos estadísticos, de los cuales se han citado algunos ejemplos anteriores, arrojan la duda por lo menos sobre el rigor de estos experimentos y a veces descreditan, incluso, la legitimidad de la validez interna reconocida de los RCT.

En cuanto a las incertidumbres que rodean la validez externa tan criticada, pensamos que son orgánicas, y a la vez la validez interna que tiene la ausencia de marco teórico las enmarca. La cuestión de la validez interna que precede forzosamente a la de la validez externa debería garantizar la -coherencia interna-, en el sentido de que los resultados de la prueba son imputables al -tratamiento- que es operado. La pérdida de control de un experimento afecta la validez interna de los resultados. Y aunque esos resultados son ampliamente reproducibles: replicabilidad y validez interna son propiedades distintas, sabiendo que una falta de replicabilidad, deja, sin embargo, presagiar una escasa validez interna. En un panorama muy esclarecedor sobre las metodologías en economía experimental, Daniel Serra describe notablemente esta tensión validez externa/ validez interna señalando: “Se apremia de inmediato la existencia de una especie de tensión entre la exigencia de validez interna y la de la validez externa, en la medida en que los diseños experimentales más aptos para conferir la primera –pasando por un nivel elevado de control son susceptibles de jugar en contra de la segunda, acentuando el carácter -insípido- del marco experimental” (Serra, 2012, p.36).

Después de haber considerado que el marco teórico es una mala guía, Banerjee y Duflo (2009: p.693) dicen más adelante en el mismo artículo, haber oído las críticas: “los experimentos están apareciendo como una herramienta poderosa para comprobar las teorías”.

No obstante, cuando se examina la cuestión relativa a la prueba de una teoría por medio de un test empírico, nunca se puede saber si el resultado es atribuible a una de las hipótesis auxiliares implícitas o no identificadas, o a la teoría probada, tanto que los datos que informan de los efectos combinados de la hipótesis que se está probando son difícilmente aislados del conjunto de hipótesis a las que son asociados según el proyecto de crítica global inspirado en el criterio Duhem-Quine.²

Crítica del análisis contrafactual

Por otra parte, las reflexiones de Smith (1990) sobre este enfoque permiten dudar de la solidez de la inferencia causal utilizada por los promotores de los experimentos aleatorios controlados: “¿Cómo el investigador puede saber que las variables convenientes se han incorporado en el estudio experimental? Del mismo modo, para la especificación de los modelos que analizan los datos obtenidos en los estudios experimentales, la teoría es la última guía.” Para él, si se mide a nivel micro, la intervención se efectúa siempre a nivel macro, y los análisis a nivel individual tienen poco interés en la acción. Dawid (2000, pp.407-448) va más allá en la crítica del enfoque contrafáctico precisando que hay dos tipos de problema de inferencia causal, ya indicados por Holland (1986): por una parte, los *efectos* de las causas consistentes en comparar las consecuencias esperadas de las distintas intervenciones posibles en un sistema, y por la otra, las *causas de los efectos*, donde se busca comprender la relación causal entre un resultado ya observado y una intervención anterior. Lo que se ilustra con la pregunta siguiente: “Tengo migraña. ¿Sería útil que tome aspirina?” Y agregar: “Mi migraña ha pasado. ¿Es porque tomé aspirina?”. Todo el problema es entonces comprender la relación causal entre un fenómeno que ya ha ocurrido y una intervención anterior. Al investigar el efecto de las causas mediante el método contrafáctico, Dawid (2000) considera entonces que el análisis se basa en una actitud que él llama *fatalismo*, para las diferentes respuestas potenciales $Y_i(u)$, cuando el tratamiento i es aplicado a la unidad u , como los atributos predeterminados de la unidad u solo esperan ser encontrados por la experiencia. También observa: “Dado que la unidad u se considera individual y no repetible, nunca hay ninguna posibilidad de probar empíricamente esta hipótesis fatalista, que puede considerarse, así como metafísica.”

La inferencia causal que resulta de los enfoques contrafactuales que incitan una inducción por eliminación, no son más que otras formas de inferencia inductiva, apenas una tentativa, y la consideración de entidades no observables debilita los fundamentos empíricos (Dawid, 2000, 2007).

Nancy Cartwright habla de un ángulo a la vez ontológico y pragmático, la exigencia ontológica que opera como garantía de la eficacia pragmática (Guillin, 2013, p.82): “Para responder a esta exigencia práctica y a esta mirada pragmática, es indispensable que las explicaciones movilizadas en los experimentos aleatorios controlados (RCT) puedan servir

² ² Citado por Daniel Serra en: “Principes méthodologiques et pratiques de l’économie expérimentale : une vue d’ensemble”, Vrin, *Revue de philosophie économique*, 2012/1, vol. 13, p. 21 à 78. Véase también Quine, W. V. O. (1951: 20-43) et Duhem, P. (1914, 2016).

de base a una intervención eficaz en lo real, es decir que proporcionen pruebas suficientes de la aplicabilidad del programa de prueba (la entrega del fertilizante justo después de la cosecha, es el ejemplo de los agricultores kenianos) fuera del contexto de la prueba experimental. Así, todo el problema es saber cómo asegurarse de esta aplicabilidad. Para Cartwright, la solución pasa por la clarificación del estatuto ontológico de las -capacidades-, de las identidades o de los agentes considerados en los RCT. En experimentos aleatorios como los practicados por el J-PAL, estas capacidades no reúnen las condiciones sólidas en la medida en que serían más bien experimentaciones que sirven para ponerlos en evidencia. Tampoco tienen la capacidad de ser irreductibles en la medida en que aún no sea posible efectuar el inventario de esas disposiciones para que una teoría unificada establezca una jerarquía de disposiciones y mecanismos en funcionamiento. Por último, carecen de leyes de composición que describirían los mecanismos equivalentes, sin hablar de composición que teorice las relaciones entre mecanismos psicológicos y no psicológicos que constituyen fenómenos sociales. Incluso cuando los RCT tienen una validez interna elevada, la metodología formal plantea limitaciones serias sobre las hipótesis que una población blanca debe respetar para justificar la exportación de una conclusión de esta hacia el objetivo.

Cartwright citada por Guillin (op.cit.) es de hecho mucho más severa porque concluye su juicio con: “En ausencia de una teoría de las capacidades de las causas que estudiamos, el valor probatorio de un RCT no solo se debilita, sino... se reduce a nada:”

Finalmente, James Heckman describe sus métodos de empirismo ciego que, no guiado por un marco teórico de la interpretación de los hechos, no conduce a ninguna parte. Para él, el enfoque Neyman-Rubin se concentra exclusivamente en los -efectos causales medios- y los efectos del tratamiento a nivel de la población para las políticas realmente aplicadas, pero no aprovisiona ningún marco para la extrapolación del descubrimiento de nuevos entornos o para las predicciones para políticas nuevas. Él agrega que, puesto que la aleatorización es utilizada para definir los parámetros de interés, a veces conduce a la confusión de que esta es la única forma o al menos la mejor para identificar los parámetros causales a partir de datos reales. La verdad, no es siempre el caso, dice.

Describe este uso de la inferencia causal contrafactual como un enfoque basado en el velo de la ignorancia retomado y ampliamente utilizado en el trabajo, por ejemplo, aplicado para evaluar diferentes distribuciones de ingresos. Los resultados podrían depender, además, de lo que las personas pueden beneficiarse de una política o también de otra. En conclusión, las hipótesis desprovistas de marco teórico implican la existencia de entidades imaginarias, cuya existencia se postula y no son fundadas epistemológicamente (R Thom, 1986), en estos casos los saltos inductivos en una situación real parecen una mala apuesta (Nancy Cartwright, 2009).

Sin atacar a las causas estructurales que generan la desigualdad, estos métodos figuran como marcadores que actúan en la lucha contra la pobreza de las políticas neoliberales cuyo consenso de Washington constituye el avatar más emblemático, pero también el fracaso más evidente. Su objetivo que se posiciona contra el enfoque macroeconómico tiene por efecto la primacía concedida a los comportamientos y tiene la racionalidad limitada al detrimento de cuestiones globales como las desigualdades, que conducen a la disolución de la cuestión del desarrollo. Pretenden así, hacer el inventario de todos los fenómenos observables para contribuir a sentar las bases de una nueva economía del desarrollo, haciendo un buen funcionamiento de la economía política no más que desembocar en proposiciones que, puesto que ellas se interesan únicamente en las capacidades individuales y no en las capacidades

colectivas, no atacan ni al proceso de exclusión ni a las causas de la marginación y mantienen en el estado las causas estructurales de la pobreza. La pretensión de los RCT en participar en un giro de la economía del desarrollo no está basada en los límites que hemos mostrado antes. Los pobres son culpabilizados, y su suerte es en su mayoría debida a las -mentalidades- inadecuadas heredadas que inducen sus decisiones erróneas. La aversión global y general explica de forma amplia los bricolajes que tienden a regenerar una economía del desarrollo (Servet, 2018).

Conclusión

La utilización sistematizada de la evaluación de las políticas públicas que comprende tanto las dimensiones normativas como las positivas, alentada, sobre todo, por las organizaciones internacionales (FMI y el Banco Mundial), los donantes y otros proveedores de fondos ha favorecido el desarrollo de un mercado de producción de saberes científicos y mediáticos que articula los sistemas de conocimientos y prácticas del gobierno con el fin de enfocarse en las prácticas de desarrollo, despolitizando el desarrollo por expertos que buscan reemplazar “el gobierno de los hombres” por “el gobierno de las cosas”, según la expresión de Dupuy (2012, p. 5).

Este artículo se ha propuesto mostrar los límites, las aporías y las fallas epistémicas, silenciadas frecuentemente, y ha contribuido en la deconstrucción de la pretensión forzada de la validez interna conferida por la eliminación de los medios de selección mostrando subsecuentemente que los problemas de generalización que conocen los RCT son, precisamente, debidos, por un lado, a la debilidad de la validez interna, y a la ausencia de teoría, por el otro.

Referencias bibliográficas

- ANGRIST J. & PISCHKE J. (2010), “The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics”, *Journal of Economic Perspectives*, pages 3 à 30.
- BANERJEE A. & DUFLO E. (2009), « L’approche expérimentale en économie du développement », *Revue d’économie politique*, 2009/56, vol. 119, pages 691 à 726.
- BANERJEE A. & DUFLO E. (2012), *Repenser la pauvreté*, le Seuil. 432 pages.
- BANERJEE, A., BARDHAN P., BASU K. & MOOKHERJEE D. (2005), *New Directions in Development Economics: Theory or Empirics? A Symposium in Economic and Political Weekly*, in *Economic and Political Weekly*, 2005, pages 4340-4344.
- BARDET F. & CUSSO R. (2012), « Les essais randomisés contrôlés, révolution des politiques du développement ? Une évaluation par la Banque mondiale de l’empowerment au Bangladesh », *Revue française de socio-économie*, 2012/2, n° 10, pages 175 à 198.

- BEDECARRATS, F., GUERIN, I. & ROUBAUD, F. (2013), « L'étalon-or des évaluations randomisées : du discours de la méthode à l'économie politique », *Sociologies pratiques*, N° 27, pp. 107 à 122.
- BURNSIDE, C. & DOLLAR, D. (1997), "Aid, Policies and Growth", The World Bank Research Department, pp. 1 à 64.
- CARTWRIGHT, N. (2009), "If No Capacities, then No Credible Worlds, But Can Models Reveal Capacities", *Erkenntnis*, 70, pp. 45-58.
- COURGEAU, D. (2000), Réflexions sur la causalité en sciences sociales, in *Recherches et Prévisions*, N°60, 2000, Chômage et famille, pp. 49-60.
- COURGEAU, D. (2012), « Peut-on parler de causalité en sciences sociales ? Les approches contrefactuelle et mécaniste », in *Les rencontres de statistique appliquée : la causalité en sciences humaines et sociales : « expliquer, modéliser, mesurer »*, Colloque conjoint entre l'INED et la Société française de statistique.
- DAWID, P. (2000), Causal inference without counterfactual, *Journal of American Statistical Association*, 95, pp. 407-448.
- DAWID, P. (2007), *Fundamentals of Statistical Causality*, RSS/EPSRC Graduate Training Program University of Sheffield.
- DEATON, A. (2010), "Instruments, Randomization, and Learning about Development", *Journal of Economic Literature*, 48, pages 424-455.
- DEATON, A. & CARTWRIGHT, N. [2018], "Understanding and Misunderstanding Randomized Controlled Trials", *Social Science & Medicine*, Elsevier. Volume 210, August 2018, pp. 2-21.
- DESROSIERES, A. (2008), *Pour une sociologie historique de la quantification. L'argument statistique*, tome I, Presses des Mines, 329 pages.
- DUFLO, E., GLENNISTER, R. & KREMER, M. (2007), "using Randomization in Development Economics Research: A toolkit", in *Handbook of Development Economics*, Elsevier, p.3895-3962.
- DUFLO, E. (2009), *Expérience, science et lutte contre la pauvreté*, Fayard, Paris, 60 pages.
- DUHEM, P. (1914), *La théorie physique, son objet, sa structure*, B/2S ENS éditions (2016).
- DEVAUX SPATARAKIS, A. (2014), « La méthode expérimentale par assignation aléatoire : un instrument de recomposition de l'interaction entre sciences sociales et action publique en France ? ». Thèse de Doctorat. Université de Bordeaux.
- DUPUY J.P (2012), *l'avenir de l'économie*, Flammarion, Coll. « Essais », 291 pages.
- ECONOMIC AND POLITICAL WEEKLY, 2005, pages 4340-4344.
- EASTERLY W. (2009), *le Fardeau de l'homme blanc : l'échec des politiques occidentales d'aide aux pays pauvres*, ed. Markus Haller.
- EASTERLY W. (2006), *les Pays pauvres sont-ils condamnés à le rester ?* Editions d'organisation, Groupe Eyrolles.
- FAULKNER, W. (2012), "Norm, Mistake, or Exemplar? A complexity Approach to IFPRI-PROGRESA in Mexico", (http://wnfaulkner.files.wordpress.com/2012/08/faulknerw_negotiatedperspectives.pdf).
- FAVEREAU J. (2016), in *Dictionnaire des conventions : autour des travaux d'Olivier Favereau*, édité par P. Batifoulie, F. Bessis, A. Ghirardello, G. de Larquier, D. Remillon, Presses universitaires du Septentrion.

- FOUGERE, D. (2010) “ Les méthodes économiques d’évaluation ”, *Revue française des affaires sociales*, 2010/1, pp. 105-128.
- FOUGERE, D. (2012) “ Les méthodes d’expérimentation en question ”, *Education et formation*, n° 81, pp. 41-47.
- FRANK R. (2007), « Peut-on accroître le pouvoir explicatif des modèles en économie? », in A. Leroux, P. Livet, *Leçons de philosophie économique*, t. III, Paris, Economica.
- FAVEREAU J. (2016), in *Dictionnaire des conventions : autour des travaux d’Olivier Favereau*, édité par P. Batifoulier, F. Bessis, A. Ghirardello, G. de Larquier, D. Remillon, Presses universitaires du Septentrion.
- GUILLIN V. (2013), « De quoi les essais contrôlés randomisés sont-ils capables ? Evaluation, mécanismes et capacités en sciences sociales », *Cahiers philosophiques*, 2013/2, n° 133, pages 79 à 102.
- HECKMAN, J. & SMITH J. (1995), “Assessing the Case for Social Experiments”, *The Journal of Economic Perspectives*, vol. 9, n° 2, *American Economic Association*, pp. 85-110
- HECKMAN J. (2005), “The Scientific Model of Causality”, *Sociological Methodology*, 35(1), pages 1 à 97. <https://doi.org/10.1111/j.0081-1750.2006.00164.x>.
- HOLLAND, P. W. (1986), « statistics and causal inference », *Journal of The American Statistical Association*, Vol 81, N 396, pp. 945-960.
- JATTEAU, A. (2013), « Expérimenter le développement ? Des économistes et leurs terrains », *Belin/Genèses*, 2013/4, n° 93, pp. 8 à 28.
- LABROUSSE, A. (2010), « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d’un outil de preuve et de gouvernement », *Revue de la régulation*, n°7.
- LABROUSSE, A. (2016), « Apprendre des expérimentations aléatoires : promesse de scientificité, complications pratiques, expériences historiques », la vie des idées (4 mars 2016).
- LEWIS, D. (1973), *Counterfactuals*, Blackwell Publishers, Harvard University Press, 156 pages.
- L’HORTY, Y. & PETIT, P. (2011), « Evaluation aléatoire et expérimentations sociales », *Revue française d’économie*, 2011/1, Volume XXVI, pp. 13 à 48.
- MAHIL, A. & TREMBLAY, D.G. (2015), « Théorie de l’acteur-réseau » in *Sciences, technologies et sociétés de A à Z*, Presses de l’Université de Montréal, pp. 234-237.
- MANSKI, C. (1997), “Monotone Treatment Response”, *Econometrica*, Vol 65, N° 6, pp. 1311-1334.
- MASSIEU, V. (2011), « L’expérimentation comme exemple de transposition d’une méthode scientifique hors de son cadre d’origine », intervention – 8^e congrès de droit constitutionnel, 2011, Nancy, (<http://www.droitconstitutionnel.org/congresNancy/comN3/MassieurT> »).pdf).
- PAULRE, P. (2007), « La causalité en économie : mises en perspective et approches opérationnelles », in *Leçons de philosophie économique. Sciences économiques et philosophie des sciences-Tome 3*, Economica, pp. 211-282.
- QUINE, W. V. O. (1951), «Two Dogmas of Empiricism», *The Philosophical Review* 60 (1951), pp. 20-43.
- QUENTIN, A. & GUERIN, I. (2013), « La randomisation à l’épreuve du terrain, l’exemple du projet de micro-assurance au Cambodge », *Revue Tiers Monde*, N° 213, pp. 179-200.

- ROUX, S. (2015), « Approches structurelles et non structurelles en micro-économétrie de l'évaluation des politiques publiques », *Revue française d'économie*, 2015/1 Vol. (XXX), pp. 13 à 65.
- RODRIK D. (2008), *The New Development Economics: We Shall Experiment, but How Shall we Learn ?* John F. Kennedy School of Government, Harvard University, RWP08-055, October.
- SACHS J. (2005), *The End of Poverty*, New York, The Penguin Press.
- SAINT-SERNIN, B., (2002), La causalité, in Andler, D., Fagot-Largeault, Saint-Sernin, B., *Philosophie des Sciences*, Tome II, Gallimard.
- SERRA, D. (2012), « Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble », *Revue de Philosophie Economique*, 13, pp. 21-78.
- SERVET, J. M. [2018], *L'économie comportementale en question*, Paris, Editions Charles-Léopold-Mayer, 208 pages.
- SUPIOT, A. (2015), *La gouvernance par les nombres*, Fayard, 512 pages.
- THOM, R. (1986), « La méthode expérimentale : un mythe des épistémologues (et des savants ?) », in *La philosophie des sciences aujourd'hui*, sous la direction de Jean Hamburger. Paris. Gauthier-Villars.
- WALLISER, B. & PROU, C. (1988), *La science économique*, Editions du Seuil, 395 pages
- ZAMORA, P.(2011), « La méthode d'évaluation aléatoire : apports et limites », *Tracés. Revue de Sciences humaines*, Hors-série 2011, À quoi servent les sciences humaines (III), pp. 175-186.