

Tendencias actuales en la evaluación de políticas públicas*

Norman Simón Rodríguez**

<https://doi.org/10.15446/ede.v28n53.75382>

Resumen

La evaluación de políticas públicas es una disciplina que tiene como objeto el examen cualitativo y cuantitativo de las decisiones tomadas por los gobiernos para resolver problemáticas sociales. Metodológica y conceptualmente, se nutre de la economía, la ciencia política, la estadística y la computación, entre otras ciencias. En este artículo se contextualizan histórica y metodológicamente las tendencias actuales en la evaluación de políticas públicas, especialmente en la evaluación de diseño y la evaluación de impacto. También se reflexiona acerca de las potencialidades de la inteligencia artificial y el *big data* para esta disciplina.

Palabras clave: políticas públicas, evaluación, desarrollo, redes bayesianas, modelación computarizada, experimento aleatorizado.

JEL: B50, I38, O20, C01, C63.

Current Trends in Public Policy Evaluation


Abstract

Policy evaluation is a discipline dedicated to the qualitative and quantitative examination of the decisions made by governments to provide solutions for pressing social issues. Its methods and concepts come from a variety of fields, such as economics, political science, statistics and computer science, among others. This paper provides the historical and methodological background of the current trends in policy evaluation, focusing on formative evaluation and impact evaluation. It also reflects on the potential applications of artificial intelligence and big data in this discipline.

Keywords: public policy, evaluation, development, bayesian networks, computer-based modelling, randomised experiment.

JEL: B50, I38, O20, C01, C63.

* **Artículo recibido:** 30 de noviembre de 2017 / **Aceptado:** 01 de junio de 2018. El artículo fue presentado en el seminario "Avances recientes en la evaluación de políticas públicas", organizado por *International Bureau of Social and Economic Research* (IBSER) en Medellín, Colombia, el 11 de noviembre de 2017. No contó con financiación.

** Magíster en Ciencias Económicas y Politólogo por la Universidad Nacional de Colombia (Bogotá, Colombia) e investigador del grupo de investigación "Gestión y Políticas Públicas Territoriales" de la Universidad Nacional de Colombia (Medellín, Colombia). Últimas publicaciones: Tres enfoques para las políticas de lucha contra la pobreza. *FORUM. Revista Departamento de Ciencia Política*, 1(2), 177-195, 2011 y *Políticas públicas en sistemas críticos: el caso latinoamericano* (compilador). Medellín: Universidad Nacional de Colombia, 2008. Correo electrónico: nsrodriguez@unal.edu.co  <https://orcid.org/0000-0003-3014-5422>

La era de los economistas ortodoxos

La evaluación de políticas públicas se puede concebir como una disciplina en la que confluyen las ciencias aplicadas, y cuyo objeto, con la investigación, es la mejora en la calidad de vida de la sociedad desde una perspectiva política, decisoria y centralizada. Con este artículo se pretende explicar con más detalle lo que esto significa¹.

Se podría afirmar que, todas las constituciones del mundo establecen unos fines esenciales para el régimen político, entre los cuales se suelen encontrar la prosperidad general, la paz y la convivencia, el imperio de la ley, el cultivo de las artes y el intelecto, entre otros, y luego prosiguen a describir el aparato estatal, cuya labor principal será la creación de los ambientes propicios para la consecución de estos fines o, incluso, la garantía de su cumplimiento. En muchos países latinoamericanos, por ejemplo, el fracaso en el alcance de los fines esenciales de la sociedad se puede atribuir jurídicamente a falencias del Estado, a tal punto que es posible que se demande como entidad abstracta por esta causa. La premisa subyacente a esta idea es que sin el Estado sería imposible, o cuando menos muy difícil, que la sociedad alcanzara la situación ideal que se propone. Teóricos como Thomas Hobbes aun llegaron a plantear un escenario más lóbrego en el que sin el Estado los sujetos entrarían en un estado de naturaleza y caos animal, algo por supuesto indeseable.

De manera que, la sociedad tiene unos fines —usualmente definidos por sus representantes políticos, según la ideología imperante al momento de redactar la constitución—, y dichos fines deben ser promovidos o garantizados por un aparato estatal, el cual puede adoptar diversas formas organizativas, siempre y cuando éstas sean coherentes con sus propios fines, so pena de sanciones contra la institucionalidad y, también, contra los funcionarios que toman las decisiones en nombre de dicha institucionalidad.

Durante los miles de años que la humanidad ha existido en la Tierra, y especialmente desde el fin de la Edad Media, en el momento en el que nació el Estado-nación, los gobiernos han intentado de una manera u otra proveerse de una ayuda sistemática en la toma de decisiones. Los Estados constitucionales modernos, en su búsqueda de una legitimidad “legal-racional” —para usar los términos de Max Weber— han intentado usar métodos también legales-rationales para mejorar la toma de decisiones, en busca de satisfacer las expectativas de la sociedad, dictadas formalmente por los fines consagrados en la constitución y manifestadas también a través del juego político —a través de las votaciones y la opinión pública, por ejemplo—. La separación de poderes fue una de las primeras innovaciones, una en la que se procuraba evitar la captura del Estado por parte del monarca y la nobleza —el Ejecutivo—, y poner dos contrapesos, en la forma del cuerpo legislativo —los congresistas, representantes de la burguesía y otros sectores poblacionales relevantes— y

¹ La división de esta conferencia en “eras”, si bien refleja a una secuencia cronológica, no pretende establecer líneas divisorias rígidas entre ellas a manera de hitos históricos, sino más bien aportar una ayuda conceptual al lector.

la rama judicial, que garantizaría la imparcialidad en la aplicación de la ley y, adicionalmente, la adhesión de todos los ciudadanos al establecimiento constitucional.

Durante la consolidación de los Estados nacionales, que se viene dando desde el siglo XVI aproximadamente, la tecnología estatal ha venido evolucionando a la par de las necesidades de los tiempos. Cuanto más poderoso se ha vuelto el Estado, más se ha exigido de éste, y más compleja se ha vuelto su labor, sobre todo en los regímenes democráticos, en los que los funcionarios de alto nivel no pueden alienarse del todo de las necesidades de la población. También, cuanto más compleja se ha vuelto la tarea, más tecnificación se ha requerido por parte de los funcionarios, lo cual se pudo evidenciar de manera patente en el siglo XX, con la fuerza que cobraron las ciencias económicas en la vida política. Dos guerras mundiales y varias recesiones demostraron que no bastaba con la formación jurídica de los funcionarios, sino que se requerían nuevos tipos de ciencias para asesorar la toma de decisiones en las altas esferas de poder. Los economistas surgieron como una clase de intelectuales que tenían unas herramientas teóricas y empíricas útiles para responder a las preguntas de los presidentes y los ministros, especialmente después de la Gran Depresión de 1929, evento que marcó el nacimiento de la macroeconomía moderna de la mano de John Keynes (Ros, 2012).

Varias dificultades, no obstante, surgieron en aquellos días. Tres de las principales fueron las debilidades teóricas de la teoría económica, la falta de herramientas para la investigación social cuantitativa y el hecho de que no todos los problemas políticos son económicos en su naturaleza. Para empezar, la teoría económica no lograba, y aún no ha logrado, conectar sus dos subteorías centrales, la microeconomía y la macroeconomía, en una sola unidad lógica internamente coherente (Cencini, 2005), lo cual llevó a que muchos de los resultados teóricos derivados a nivel de individuos no aplicaran a nivel de la sociedad, lo cual dificultaba su adopción en políticas públicas. Esto sin contar que algunos de los supuestos tradicionales de los economistas, tales como la existencia de información perfecta —típica de la teoría de juegos—, la competencia perfecta, la racionalidad de los agentes y la homogeneidad de los agentes —que hay un agente representativo que resume perfectamente a todos los demás—, adolecían de una gran falta de realismo.

El segundo gran problema que enfrentaron los economistas y en el cual, cabe decir, tuvieron mucho más éxito, fue el de la falta de herramientas cuantitativas empíricas para testear hipótesis y teorías sobre las políticas públicas. Los estadísticos matemáticos que siguieron la tradición de Karl Pearson, entre ellos Ronald Fisher, sentaron las bases de la práctica actual de la estadística de corte frecuentista, la cual busca “dejar hablar a los datos” con la menor cantidad de suposiciones subjetivas posible. Es en la primera mitad del siglo XX que se desarrollan técnicas como los tests de significancia, los intervalos de confianza y el testeo de hipótesis, que rápidamente desbancaron a los métodos más subjetivistas de la tradición bayesiana, que ve a la probabilidad como una medida de la incertidumbre psicológica. Los economistas dedicados al estudio de las decisiones macroeconómicas y aun a los estudios

de las guerras exteriores de Estados Unidos fueron, en su vasta mayoría, entrenados en la tradición frecuentista².

Durante el siglo XX hubo una creciente exigencia por parte de los gobiernos para comprobar la efectividad de sus intervenciones, tanto a nivel técnico como a nivel político. Se buscaba saber si una política dada iba a funcionar o no —o si había funcionado en el pasado, o si estaba funcionando en el presente, dependiendo del horizonte temporal escogido—, si era la mejor alternativa posible —con análisis de costo/efectividad, entre otros— y si era políticamente conveniente. Las dos primeras preguntas son típicas de la evaluación de políticas públicas y son de orden *causal*, mientras que la última lo es del análisis de políticas públicas (Salazar, 2009).

En principio, la estadística convencional no contempla responder preguntas de orden causal, como las que se requieren en el estudio de las políticas públicas. Mientras un alcalde preguntaría: “¿cuánto se reduciría la pobreza rural si implementara una política de subsidios?”, que es una cuestión causal, el estadístico respondería, después de haber implementado los subsidios: “un subsidio de 100 dólares está asociado a una reducción de 4 puntos porcentuales en la pobreza rural”, mas no podría decir si dicha asociación es causal o no. Dicho de otro modo, no podría determinar si fueron los subsidios los que redujeron la pobreza, o si hubo una tercera variable —llamada “factor de confusión”— que causó tanto el subsidio como la reducción en la pobreza, como por ejemplo la llegada de una gran fábrica a la ciudad, la cual hizo cabildeo para que el Gobierno entregara subsidios en su localidad y, a la vez, contrató personas para trabajar en sus instalaciones. El estadístico tampoco puede decir la magnitud del impacto de la política en el indicador de desarrollo, porque puede ser que una parte de los cuatro puntos porcentuales en reducción de la pobreza sí se debiera a la política, pero que otra parte fuera causada por factores externos, tales como programas de alimentación de ONG. A la parte del efecto debida a factores externos se le denomina sesgo de variable omitida.

Esta divergencia entre las necesidades de los economistas y la oferta de los estadísticos dio origen a una rama de la estadística conocida como econometría, que es un intento por usar los elementos de la estadística matemática para responder preguntas causales. Los métodos econométricos son la esencia de la evaluación de impacto de políticas públicas, aún cuando no se usen para abordar cuestiones puramente económicas. El nombre es una herencia de la historia de la evaluación de políticas públicas.

En el centro de la econometría está el concepto de *contrafactual*, que había sido ya avisado por David Hume y adoptado por los frecuentistas para el método experimental (Fisher, 1937). Un ejemplo, en el caso de que alguien sea designado como ministro de educación, se enfrentaría a la pregunta de si debe el Estado abrir una nueva sede de una universidad pública o si debe, alternativamente, ampliar la capacidad instalada de las sedes actuales. ¿Cómo elegir? En un mundo ideal, antes de tomar la decisión se deberían tener todos los

² Orloff y Bloom (2014) para una comparación intuitiva entre la tradición frecuentista y la bayesiana.

datos con la mayor certeza posible, en particular, las respuestas a las siguientes tres preguntas: ¿qué pasaría si se abriera la nueva sede? ¿Qué pasaría si se ampliaran las sedes actuales? y, claro, ¿qué escenario es mejor? El gran problema es que no se pueden tener todos los datos al mismo tiempo. Se podría tomar la decisión de ampliar las sedes y ver qué pasa, pero eso significaría que nunca se podría saber que pasaría con la construcción de una sede nueva. O se podría crear la sede nueva y ver qué pasa, pero entonces renunciarían a saber los efectos de ampliar las sedes. Con el ejemplo que se describió se puede aducir que, aquello que se observó se denomina *factual*, mientras aquello a lo que se renunció —lo que no se puede observar—, se denomina *contrafactual*.

Se podría suponer que la decisión fue ampliar las sedes. Entonces, se responden las anteriores tres preguntas:

1. *¿Qué pasaría si se abriera la nueva sede?* No se sabría, porque es un escenario contrafactual.
2. *¿Qué pasaría si se ampliaran las sedes actuales?* Se sabría, porque es un escenario factual. Basta con medir los indicadores de interés —por ejemplo, la tasa de matrícula en educación superior— antes y después de la intervención para responder esta pregunta.
3. *¿Cuál escenario es mejor?* Al no tenerse información sino de un solo escenario, la pregunta es irrespondible.

Claramente, como ministros de educación no se sentirían muy satisfechos si contrataran un equipo de economistas que cobran grandes sumas por negarse a contestar sus preguntas. Es por esto que la econometría busca resolver la situación de la mejor manera posible y tratar de estimar el escenario contrafactual, es decir, imaginarlo. Como buenos frecuentistas, no obstante, los econométricos consideraban que dicha imaginación debía ser lo menos subjetiva posible y por eso desarrollaron un conjunto de métodos que usan las matemáticas para construir estos contrafactuales artificiales.

Durante la segunda mitad del siglo XX, el método más conocido y usado fue la regresión lineal múltiple. No era un método nuevo, de hecho su derivación matemática se la debemos a Gauss, quien la describió en el siglo XIX. Lo nuevo fue la interpretación causal que se buscó darle, y que la propulsó como método fundamental para la evaluación de políticas. Los politólogos, cuya disciplina nació por aquellos mismos días, se embarcaron en el entusiasmo de la regresión y empezaron también a explorar la econometría. Por muchos años la regresión no pudo ser destronada, aún a pesar de ser un método relativamente pobre, quizás por su facilidad de uso. Los otros métodos econométricos, en particular el método experimental de Fischer y el emparejamiento estadístico siguieron existiendo, aunque no eran tan populares.

La mención a los politólogos nos lleva a la tercera dificultad que tuvo la evaluación de políticas públicas con el énfasis de los economistas. Evidentemente, no todos los problemas sociales pueden ser modelados utilizando la metáfora de los mercados y los actores racionales, esto porque no todos los problemas sociales son de orden económico, y algunos

políticos progresistas lo habían empezado a notar. Aunque la caída del muro de Berlín dejó muy bien parados a los economistas de la corriente ortodoxa, tanto que hasta proclamaron el fin de la historia (Fukuyama, 1992), su predominio en la política macroeconómica no se trasladó necesariamente a la política social, la cual hacia la década de los noventa ya estaba buscando herramientas propias. Otras políticas públicas como las políticas de transporte, medioambiente y planeación urbana también habían estado usando subrepticiamente métodos traídos de la ingeniería.

El cambio de paradigma más marcado en la evaluación de políticas públicas de las últimas décadas ocurrió a finales de los noventa y principios de los dos mil. Entramos en la era de los economistas heterodoxos.

La era de los economistas heterodoxos

La economía tradicional, se ha dicho, hacía uso de métodos econométricos que no podían eliminar bien el sesgo de las variables omitidas, es decir, que no podían imaginar un contrafactual de buena calidad. La regresión múltiple, así como otros métodos menos sofisticados tales como las series de tiempo interrumpidas, el pre-post, las diferencias simples y las diferencias en diferencias, intentaban eliminar el sesgo controlando covariables escogidas *ad hoc*, o incluso suponiendo que no existía. Las últimas dos décadas han visto el resurgimiento del método experimental —los ensayos controlados aleatorizados— como la tecnología de punta en la evaluación de impacto. Un ensayo aleatorizado consiste, en su definición más básica, en asignar al azar un grupo de personas para que reciban la intervención o “tratamiento”, y otro grupo de personas, también al azar, para que no lo reciban. Al grupo que recibe tratamiento se le llama grupo de tratamiento, y al que no, se le llama grupo de control.

Los principales promotores de esta nueva metodología han sido los economistas heterodoxos de Massachusetts Institute of Technology, en particular la francesa Esther Duflo, entre otros³. Las ONG Abdul Jameel Poverty Action Lab —J-PAL, co-fundado por Duflo—, Innovations for Poverty Action (IPA) y 3ie han sido motores visibles de esta explosión de ensayos controlados aleatorizados en todo el mundo. En Latinoamérica y Colombia se han conducido ya varios de estos experimentos, y dichas ONG tienen sedes en América Latina —una de ellas, IPA, tiene sede en Bogotá—. Algunos politólogos, entre los que se cuentan Benjamin Olken y el beninés Leonhard Wantchekon, se pusieron a la tarea de aplicar este método econométrico a problemas eminentemente políticos, tales como la corrupción y el clientelismo en las elecciones. Pero el campo de acción más rico de este método ha sido por mucho la política social.

La forma en la que se aplica un ensayo controlado aleatorizado es bastante simple. Supóngase que se quiere evaluar el impacto de regalar libros en inglés a estudiantes de colegio. La hipótesis de la secretaria de educación es que gracias a los libros, el nivel de inglés de los estudiantes va a mejorar. Para poder comprobar esta hipótesis, como ya se ha mencio-

³ Ver por ejemplo, el excelente volumen de Duflo y Banerjee (2017).

nado, se debe estimar el escenario contrafactual, es decir, habiendo regalado los libros hay que averiguar qué habría pasado si no hubieran sido regalados. Los investigadores simplemente hacen un sorteo —posiblemente secreto— en el que participan todos los salones o cursos de bachillerato de todos los colegios de la ciudad; supóngase que son 500 salones. A partir de esta asignación aleatoria sale un listado de 250 salones a los que se les deberá dar libros, y otros 250 a los que no —estos últimos son el grupo de control—. Lo que sigue después es repartir los libros en el grupo de tratamiento y esperar unos meses a que la intervención haga efecto. Entre tanto, se va haciendo un seguimiento a ambos grupos. Al cabo del tiempo de espera se evalúa el nivel de inglés de los estudiantes y se compara el nivel del grupo de tratamiento con el del grupo de control.

Para efectos de la explicación, se puede suponer que las tres preguntas tienen las siguientes respuestas:

1. ¿Qué habría pasado si se hubieran regalado los libros? El nivel de inglés en el grupo de tratamiento mejoró diez puntos porcentuales.
2. ¿Qué habría pasado si no se hubieran regalado los libros? El nivel de inglés en el grupo de control mejoró siete puntos porcentuales.
3. ¿Qué escenario es mejor? El escenario mejor es en el que se regalan los libros, pues la diferencia entre ambos escenarios fue 3pp. A dicha diferencia se le denomina *impacto*, y es precisamente el número al que se busca llegar —en rigor, esta cifra no es el impacto sino una estimación del impacto, ya que el impacto real no se puede conocer, puesto que el contrafactual teórico no existe—.

Lo que garantiza que no hay sesgo es la asignación aleatoria, porque ambos grupos eran iguales en promedio antes de la intervención.

Los ensayos aleatorizados que se están realizando en el mundo hasta la fecha se cuentan por los cientos, y de todos ellos se está extrayendo evidencia valiosa para informar las decisiones políticas de los funcionarios del Estado. Poco a poco, los economistas heterodoxos han estado desplazando a los ortodoxos en la evaluación de políticas y programas, tal como lo demuestra el hecho de que, por ejemplo, el mismo Banco Mundial ya esté realizando publicaciones masivas de economistas heterodoxos dedicados a la evaluación de impacto (Gertler, 2017).

El predominio de los “randomistas”, como se han denominado estos economistas, no ha impedido que se sigan dando avances en otros métodos econométricos. Desde el 2012, por ejemplo, se ha estado describiendo un nuevo método de emparejamiento estadístico que puede llegar a reemplazar el que ha dominado por cerca de cuarenta años. Se trata del emparejamiento exacto por redondeo —*Coarsened Exact Matching* (CEM)—, propuesto por Gary King y otros (Lacus, King y Porro, 2012), el cual, a diferencia del tradicional emparejamiento por puntajes de propensión (PSM por sus iniciales en inglés), no resume las covariables en un número para emparejar las unidades del grupo de tratamiento con unidades del grupo de control, sino que realiza el emparejamiento al mismo tiempo en todas las

covariables sin resumirlas, sólo redondeándolas o truncándolas. La lógica de los emparejamientos es similar a la de los experimentos aleatorizados, puesto que en ambos métodos se busca crear un grupo de control que sea exactamente igual al grupo de tratamiento, en promedio, para que sea un contrafactual válido. La diferencia radica en que mientras en el experimento dicho grupo de control se construye usando la asignación aleatoria, en los emparejamientos este se construye buscando un “gemelo” para cada uno de los elementos del grupo de tratamiento. Se consideran “gemelos” dos individuos que sean iguales o parecidos en un conjunto de características, las covariables, como por ejemplo la estatura, el sexo y el nivel educativo. En el PSM las covariables se resumen en un número, llamado puntaje de propensión, y si un individuo tiene un puntaje de propensión parecido al de otro que está en el grupo de tratamiento, se le empareja con este y se le pone en el grupo de control. En el CEM, las covariables se redondean o se truncan y luego se hace el emparejamiento entre individuos que tengan exactamente los mismos valores truncados. Para ilustrarlo con un ejemplo, supóngase que la única covariable de interés es “horas trabajadas al día”, de modo que el objetivo es buscar dos individuos que tengan la misma cantidad de horas trabajadas al día para emparejarlos. El individuo X trabaja 7.7 horas diarias, y el individuo Y trabaja 8.4 horas diarias. Si se redondean ambos números, los dos quedan en 8, y pueden por tanto ser emparejados. Los ensayos controlados aleatorizados son superiores a los emparejamientos, pero cuando no es posible hacer un ensayo aleatorizado, el CEM es generalmente preferible al PSM (King y Nielsen, 2016).

La era de los profesionales eclécticos

La evaluación de políticas públicas como disciplina no ha dejado de evolucionar, claro está, y se pueden avizorar nuevas tendencias las cuales, paradójicamente, parecen viejas y nuevas al mismo tiempo. Al inicio de este artículo se dijo que la evaluación de políticas públicas se puede concebir como una disciplina investigativa en la que confluyen las ciencias aplicadas, y cuyo objeto es la mejora en la calidad de vida de la sociedad desde una perspectiva política, decisoria y centralizada.

Que la perspectiva es política, decisoria y centralizada queda claro a partir de todo lo que se ha hablado hasta el momento. Se busca responder a las preguntas de los funcionarios del Estado, quienes responden a presiones políticas; se busca asesorar la toma de decisiones; y dicha toma de decisiones es realizada por funcionarios que ejercen el poder de manera centralizada, esto es, por personas que tienen la capacidad de exigirles a otros miembros de la sociedad la obediencia a sus determinaciones.

Para comprender la diferencia entre las decisiones centralizadas y las descentralizadas, puede resultar útil una analogía de las ciencias biológicas. En un cardumen no hay ningún pez que sea líder y que esté decidiendo adónde virar, ni tampoco un pez que comunica a los demás hacia dónde deben virar. No obstante, los movimientos de los cardúmenes bajo el agua son asombrosamente sofisticados y coordinados. A este tipo de coordinación “espontánea”, sin líder, se le llama coordinación descentralizada, muy común en insectos sociales como las hormigas (Ge *et al.*, 2012). En contraste, los sistemas en los que hay liderazgos y

comunicación de órdenes o instrucciones, son sistemas centralizados. Así, pues, la evaluación de políticas públicas tiene una perspectiva centralizada en tanto busca influenciar a los tomadores de decisiones para que éstos, a su vez, influyeran a través del poder político a los sujetos que componen la sociedad.

De este modo, el evaluador de políticas públicas se distancia en parte de los investigadores sociales que buscan estimular a la sociedad para que esta se autoorganice, y de los que postulan críticas al sistema político en general, porque su labor es menos ambiciosa. El evaluador no trata de entender las falencias del sistema político ni tiene la expectativa de una sociedad autoorganizada, sino que parte de la realidad de que hay un sistema político, por defectuoso que pueda ser, y trata de incidir en la toma de decisiones de los que detentan el poder, quienes pueden organizar a la sociedad centralizadamente.

El objeto del evaluador, sin embargo, no deja de ser moralista en el sentido de que se busca mejorar la calidad de vida de las personas, tanto a un nivel material —monetario— como a un nivel psicológico. Este es otro de los componentes de la definición. Y también se dijo que cuando se evalúan políticas se está haciendo uso de la investigación y de las ciencias aplicadas. Esto significa que hay un interés práctico en toda evaluación de políticas, se busca resolver un problema o aprovechar una oportunidad, aunque también hay un interés académico. Estas ciencias aplicadas no se restringen a las ciencias sociales, puesto que incluyen a también las ciencias duras.

Más adelante se hará referencia a algunos horizontes investigativos que potencialmente marcarán las tendencias de la evaluación de políticas públicas en el futuro. Se pueden clasificar en subcategorías, según el tipo de evaluación con el que más se relacionan⁴:

1. Evaluación de impacto: experimentos bayesianos, generalizabilidad.
2. Evaluación de diseño: redes bayesianas, simulaciones de Monte Carlo, dinámica de sistemas, modelación basada en agentes.
3. Campos prometedores: *big data*, aprendizaje automatizado —*machine learning*—.

Casi todas estas herramientas se dan a partir de desarrollos antiguos, o han sido redescubiertas recientemente.

Evaluación de impacto

Como ya se ha señalado, la necesidad primordial del evaluador de impacto es construir contrafactuales convincentes para poder dar respuesta a sus tres preguntas de investigación. A

⁴ Se puede clasificar la evaluación de políticas públicas en cuatro clases según su alcance: la evaluación de impacto mide los cambios en la calidad de vida de la población que son atribuibles a la intervención; la evaluación de resultados reporta las acciones realizadas por el Estado en el marco de la intervención; la evaluación de implementación analiza la puesta en marcha del programa a nivel burocrático, administrativo y contractual; y la evaluación de diseño —la cual es prerequisite para todas las demás— estudia la coherencia interna del programa y qué tan razonablemente sus supuestos le permitirán causar el impacto —a esta cadena de causalidades y supuestos se la llama convencionalmente “teoría del cambio”—.

principios del siglo pasado los frequentistas alcanzaron mayor popularidad que los bayesianos; la tendencia actual, aparentemente, es que cada vez más los bayesianos se acercan a cooperar con los frequentistas en igualdad de condiciones y a estar de moda de nuevo. Por ejemplo, la FDA —Departamento Administrativo de Alimentos y Medicamentos de Estados Unidos— ha empezado a aceptar ensayos aleatorizados bayesianos para aprobar nuevos tratamientos y dispositivos médicos (Fienberg, 2011).

Un experimento frequentista toma el grupo de tratamiento y el grupo de control como dos muestras equivalentes de una población, y estima la diferencia post-tratamiento entre ambos grupos sobre la suposición —imaginaria— de que si se repitiera ese proceso millones de veces, dicha diferencia estimada reflejaría con cierto nivel de confianza la diferencia real entre lo que le habría pasado a la población con el tratamiento y lo que le habría pasado a la población sin el tratamiento. En otras palabras, con muestreos repetidos infinitamente, el estimativo de impacto —la diferencia entre el grupo de tratamiento y control— se acercaría en promedio al impacto real —la diferencia entre la población con tratamiento y la misma población sin tratamiento, si ambas cosas pudieran suceder al mismo tiempo—.

El experimento bayesiano, por su parte, comienza con la postulación de una distribución de probabilidad a priori que representa lo que los investigadores conocen acerca del fenómeno a estudiar. Por ejemplo, la distribución a priori del efecto de las transferencias monetarias condicionadas —subsidios condicionados— en la nutrición infantil. Se puede partir bien sea de la opinión de expertos o de estudios empíricos para tratar de estimar la distribución a priori. Se puede establecer, por ejemplo, que las transferencias mejoran en promedio la nutrición infantil en una desviación estándar. Después se realiza el ensayo aleatorizado de la manera usual —aunque también se pueden hacer modificaciones al diseño, como en los ensayos adaptativos—, y luego se actualiza la distribución a priori con los datos arrojados por el ensayo. Las conclusiones del estudio se presentan como una distribución a posteriori (Simon, 1999; Spiegelhalter y Freedman, 1994). La ventaja de los ensayos bayesianos es que permiten explícitamente actualizar las creencias de la comunidad científica de una manera rigurosa y matemáticamente formalizada, pero sus probabilidades a priori son a veces difíciles de estimar; esto, sin embargo, se ve compensado por el hecho de que es posible incluir información cualitativa en las distribuciones a priori, lo cual no es práctica común en los ensayos frequentistas (Lee y Chu, 2012). El uso de ensayos controlados aleatorizados bayesianos para evaluar políticas sociales es aún incipiente, puesto que se restringe mayormente al ámbito clínico, aunque la metodología no ha sido del todo ignorada en la disciplina (Michaloupoulos, 2012). El método fue propuesto con fuerza en los noventa, pero aún no ha sido adoptado por los evaluadores de políticas públicas de manera masiva⁵.

Paralelamente al problema de la creación de contrafactuals se encuentra el problema de la generalizabilidad. Si, por ejemplo, se hace un ensayo aleatorizado en Botsuana y se encuentra que gracias a un programa educativo para padres, los niños asistieron más

⁵ Para una introducción amigable a los ensayos bayesianos adaptativos, consultar Berry (2006).

a la escuela, ¿podría esa misma conclusión extrapolarse a una aldea en Bolivia, o a una ciudad colombiana?

A la capacidad de generalizar un resultado se le llama “validez externa”. El gran problema de la validez externa es que es prácticamente imposible de conseguir, lo cual implicaría que ese mismo programa debería ser testeado en Bolivia, y en Colombia, y en cada lugar donde se quisiera aplicar, una vez tras otra, sin la posibilidad de aprender de las conclusiones de los estudios anteriores. La razón para que la validez externa sea tan elusiva proviene del hecho de que las muestras con las que se hacen los estudios aleatorizados son representativas únicamente de su población particular, no de otras poblaciones. Y lo que es más desalentador, algunas veces ni siquiera son representativas de su propia población, porque con el paso del tiempo, es posible que la muestra y la población vayan presentando diferencias sistemáticas entre ellas debido, por ejemplo, a que el Estado decida usar esa misma muestra para aplicar otras intervenciones no contempladas en la investigación —como regla general, cuanto más dure el ensayo, más peligro hay de que la muestra pierda representatividad—.

Cartwright y Munro (2010) mostraron además que este problema no se debe sólo a la falta de representatividad de la muestra en relación con la metapoblación —o la población—, sino también a la falta de una predefinición teórica de los mecanismos causales que lleven a los efectos en los ensayos aleatorizados. En otras palabras, para establecer una generalización relativamente aceptable, se hace estrictamente necesario contemplar la *teoría del cambio* de la intervención (Cartwright, 2014). Frente a la crítica de estas autoras y de otros investigadores (Basu, 2013; Pritchett y Sandfur, 2013), los randomistas, sobre todo en 2017, han estado haciendo un énfasis mayor en la creación de teorías del cambio para las evaluaciones de impacto, y han estado formulando directivas esenciales para la generalizabilidad de las mismas.

El marco conceptual para la generalizabilidad que se ha venido proponiendo (Bates y Glennerster, 2017; Glennerster, 2017) se puede sintetizar en tres partes. La primera corresponde a las condiciones locales del lugar donde se está implementando la política, las cuales no se pueden generalizar. Por ejemplo, la edad a la que normalmente se independizan los jóvenes en un país puede ser muy diferente a la de otro país, y eso, necesariamente, tendrá un impacto en los prospectos económicos de los empleados jóvenes; un estudio hecho en Estados Unidos, donde dicha edad es baja, no se puede extrapolar automáticamente a un país latinoamericano en donde esa misma edad sea alta. La tercera parte —más adelante se mencionará la segunda— también es de índole local, más no se refiere a las precondiciones existentes antes de la intervención, sino a los eventos particulares que surgieron durante la implementación del programa o de la evaluación como tal. Esta parte, que es lo que normalmente se estudia cuando se hace evaluación de implementación, tampoco se generaliza.

La segunda parte del marco de Bates y Glennerster corresponde a las “teorías sobre el comportamiento humano”. Así, si por ejemplo en un ensayo aleatorizado se les mostró a unos padres de familia que los ingresos esperados de sus hijos mejorarían si éstos estudiaran y

los padres los enviarán más a la escuela, esto podría ser indicativo de que, *en general*, las decisiones educativas de los seres humanos pueden responder fuertemente a incentivos monetarios futuros. Aquí la economía del desarrollo se empieza a solapar con la psicología, la sociología y la ciencia política. Si el *zeitgeist* de la disciplina en el siglo XX respondió al *ethos* de la economía ortodoxa y el de las últimas décadas ha sido marcado por los economistas heterodoxos, es posible que el de los próximos años lo sea por los científicos sociales, más no por los que se dedican a una sola ciencia social, sino por los que se dedican eclécticamente a todas, e incluso a otras ciencias no sociales. En todas las tres fases de este marco conceptual para la generalizabilidad el rol de la intuición es clave (Basu, 2013).

Evaluación de diseño

La evaluación de diseño es la evaluación de la teoría del cambio de las intervenciones y, a diferencia de la evaluación de impacto, sí puede hacerse antes de la implementación de las políticas o programas, lo cual la hace muy atractiva para tratar de estimar el éxito de las políticas *ex-ante*. En este aspecto que se puede considerar que la evaluación de diseño tiene un campo de acción muy prometedor, en especial en lo concerniente a la simulación computarizada de políticas públicas. Se propone pensar por un minuto en un juego de fútbol. En la vida real, un juego de fútbol es impredecible, no sólo porque el marcador es impredecible, sino también porque las jugadas de cada uno de los veintidós jugadores son absolutamente impredecibles. Para usar un lenguaje más abstracto, un juego de fútbol es impredecible a nivel de impactos y también lo es a nivel de procesos. Si fuera totalmente predecible, no sería para nada emocionante ver fútbol. Sin embargo, cualquiera estará de acuerdo en que a pesar de que no es posible saber cómo se desenvolverá un partido de antemano, sí es posible encontrar patrones en los partidos de fútbol y tratar de hacerse a una idea —buena aunque no perfecta— de lo que pasará en un partido dado. Por decir, si se pusiera a jugar al Barcelona contra un equipo local de Colombia, para la mayoría de la gente sería muy fácil apostar a que el Barcelona ganaría, y a que Lionel Messi haría algún tipo de jugada divertida en el proceso. Habría posibilidades de que el Barcelona perdiera, claro, pero serían más bien remotas.

Desde el punto de vista del director técnico del Barcelona —que es un punto de vista político, decisorio y centralizado—, es necesario identificar aquellas acciones que, si se hicieran, darían una ventaja sostenible al Barcelona por encima de cualquier otro equipo del mundo por el mayor tiempo posible. Es imposible predecir cómo se desempeñará el equipo en cualquier partido dado, pero se puede aprender de la experiencia para detectar lecciones como que hay que fomentar el trabajo en equipo, que hay que contratar *cracks*, que hay que maximizar el tiempo de retención de la pelota, que hay que promover la disciplina futbolística y personal entre los jugadores, etcétera. Si se puede pensar razonablemente que estas decisiones causan el desempeño satisfactorio, el técnico muy seguramente las va a implementar.

Esto puede funcionar en el mundo del fútbol, en donde ganar o perder un partido no es cuestión de vida o muerte, y en donde hay partidos prácticamente todos los días del año, de modo que se puede ir formando una “política ideal” a partir del ensayo y error. En el mundo

social las cosas son diferentes, porque el fracaso de una política pública sí puede conllevar a consecuencias muy negativas desde todo punto de vista, y porque usualmente no se tiene el lujo de aprender por ensayo y error en repetidas versiones de la intervención. ¿Qué se puede hacer, entonces, para encontrar “políticas ideales”?

En este mismo sentido, se propone pensar en un videojuego, como por ejemplo el videojuego oficial de la FIFA. En un videojuego suficientemente realista, el director técnico del Barcelona podría probar diferentes formaciones tácticas, diferentes nóminas, diferentes niveles de trabajo en equipo, etcétera, y luego decidir la combinación óptima de medidas, la que incrementó la probabilidad de tener impactos grandes —marcadores ventajosos— después de haber jugado muchos partidos en realidad virtual. En otras palabras, el director técnico podría detectar los patrones que llevarán al éxito a su equipo sin necesidad de realizar partidos en la vida real —con los costos y riesgos que ello implica—, sino sólo jugando en su consola.

Algo similar se puede hacer con las políticas públicas, aplicando modelos de simulación computarizada (Gentile, Glazner y Koehler, 2015). La idea es tan simple como que el tomador de decisiones, junto con los demás actores interesados en la política, incluyendo a la ciudadanía, pueden hacer “juegos” en realidad virtual en los que ya no se está jugando un partido de fútbol, sino que se está en una sociedad o comunidad virtual. Hay varios tipos de modelación, entre ellos la modelación con ecuaciones diferenciales —que incluye la dinámica de sistemas— y la modelación basada en agentes. Cada una de las dos proviene de una tradición diferente, pero puede ser adoptada con facilidad en el ámbito de las políticas públicas. Lo que sorprende es que la adopción sea tan lenta —esa es la queja de Ghaffarzadegan, Lyneis y Richardson (2011)—, y por eso, aunque se puede pensar en ellos como “avances” o “tendencias” recientes, a la verdad son más bien gemas escondidas por muchos años que necesitan ser redescubiertas.

La modelación basada en ecuaciones diferenciales no es para nada nueva, de hecho hace parte de la caja de herramientas estereotípica de los economistas ortodoxos. Tan temprano como 1928, Frank Ramsey ya estaba creando modelos microeconómicos de equilibrio que consideraban la utilidad de los agentes y sus cantidades consumidas, y dilucidaban, según la teoría marginalista, los comportamientos de ahorro de las personas a nivel agregado (Ramsey, 1928). Modelos teóricos de este tipo siguen siendo usados rutinariamente para hacer evaluaciones de políticas, en conjunto con métodos econométricos empíricos (Todd y Wolpin, 2008). Puede considerarse que ha habido cierta dificultad para trasladar esta metodología desde el campo de lo económico al campo de las políticas públicas en general, aunque una versión de la misma, la dinámica de sistemas, promovida desde las ingenierías, ha tenido avances esperanzadores.

Lo que la dinámica de sistemas hace es, en un nivel abstracto, tomar la teoría del cambio de una política pública y definir entre sus componentes unas relaciones matemáticas expresadas como ecuaciones diferenciales que consideran explícitamente al tiempo. Son sistemas “dinámicos” precisamente porque los valores de sus variables cambian con el tiempo. La

metodología fue formulada primero en la década de 1960 (Forrester, 1969; Forrester, 1971; Meadows, Meadows, Randers y Behrens, 1972), en el contexto de la planeación urbana, y con los años se ha ido extendiendo a otros contextos, aunque no de una manera tan rápida. En parte, es posible que la dinámica de sistemas no se haya popularizado porque no fueron los economistas sus promotores, sino por lo general los ingenieros y, por otro lado, porque sus predicciones no siempre son realistas.

Esta última crítica, podría estarse debilitando en el futuro cercano a medida que se vayan popularizando los métodos estocásticos de simulación, en particular las simulaciones de Monte Carlo. Las simulaciones de Monte Carlo, que también existen desde hace años, son el siguiente paso lógico para la dinámica de sistemas en la evaluación de políticas públicas. En vez de tomar valores únicos, los elementos de la teoría del cambio toman distribuciones de probabilidad, de las cuales se sacan valores de manera aleatoria en cada simulación; el método calcula para cada simulación los valores de las variables de impacto según las ecuaciones diferenciales del modelo a lo largo de un rango de tiempo dado, y finalmente reporta los impactos simulados en forma de distribuciones de probabilidad. Paquetes de software como Vensim y Stella permiten este tipo de simulación, y otros más sencillos como InsightMaker tienen funciones que pueden ser usadas para lo mismo. Los ingenieros militares notaron la utilidad de combinar dinámicas de sistemas y simulaciones de Monte Carlo desde los noventa (Haggenson, 1990), y poco a poco la idea ha ido tomando fuerza entre los econométristas (Keane y Wolpin, 1994; Lee, 2005).

Las simulaciones de Monte Carlo, no obstante, trascienden el reino de la dinámica de sistemas y pueden ser usadas, más generalmente, para evaluar teorías del cambio cuyas relaciones no están representadas con ecuaciones diferenciales sino con diagramas de causalidad simples, que pueden tomar cualquier forma funcional matemática (Pearl, 2003). A estos diagramas de causalidad se les denomina “redes causales bayesianas”, y tienen un gran potencial para la evaluación de políticas públicas. Éstas implican un cambio notacional (Pearl, 2003), no requieren suposiciones fuertes como el equilibrio general o la racionalidad de un agente representativo, y pueden usar fácilmente evidencia cualitativa desde un punto de vista bayesiano. El cambio notacional es un tema altamente técnico, pero básicamente significa que las redes bayesianas tienen el potencial de cambiar el lenguaje matemático de la econometría de manera radical (Pearl, 2001), haciéndola mucho más accesible para los hacedores de política pública y para los investigadores de a pie, a la vez que la puede hacer más abierta a aceptar formulaciones bayesianas, tales como los ensayos aleatorizados bayesianos anteriormente mencionados.

Las redes bayesianas no requieren tampoco la suposición de que todos los individuos de la sociedad son iguales, y esa es una ventana de oportunidades muy grande. Los modelos basados en agentes —ABM, por sus iniciales en inglés— permiten, precisamente, modelar agentes heterogéneos, de modo que ya no es necesario suponer la existencia de agentes representativos, lo cual les dota de un mayor realismo frente a la modelación tradicional. Los ABM posibilitan la identificación de comportamientos emergentes y efectos de heterogeneidad. El ejemplo clásico de los comportamientos emergentes proviene del modelo

de segregación de Schelling (1978), en el que un grupo de ciudadanos, cada uno de ellos representado por una ficha en un tablero, busca tener por vecinos a ciudadanos del mismo color —en el modelo hay dos colores—; Schelling mostró que aún cuando *todos* los ciudadanos individualmente considerados tengan altas tasas de tolerancia a vecinos de otro color, *en conjunto* la sociedad tiene tasas de tolerancia más bajas, lo cual implica que, dentro del modelo, individuos tolerantes llevan a sociedades segregadas. Un comportamiento emergente es una propiedad agregada que surge a partir de la interacción entre individuos, propiedad que es considerada sorprendente o contradictoria. En la naturaleza hay muchos comportamientos emergentes, como en el caso de las hormigas que crean complejas colonias sin tener un mando central, sino sólo a partir de interacciones entre hormigas individuales que operan según reglas muy simples.

En cuanto a los efectos de la heterogeneidad, dentro de los modelos se pueden definir reglas de interacción diferenciadas entre agentes distintos. Por ejemplo, en un modelo de seguridad ciudadana, se puede definir que los policías capturen a los ladrones pero no a los ciudadanos, para ver qué comportamiento emergente se ve de estas interacciones; también se podrían establecer variaciones a esas reglas, permitiendo que los policías también puedan capturar ciudadanos cuando éstos adquieran el estatus de “cómplices”, y ver si los comportamientos emergentes son distintos. Los ABM pueden ser usados para evaluar desde el diseño de mecanismos sociales para proteger a los ancianos (Kennedy, Ihara, Tompkins y Wolf-Branigin, 2015) hasta intervenciones en salud pública para combatir la obesidad (Giabbanelli y Crutzen, 2017). Introducciones a la técnica y al software NetLogo, el más popular para este tipo de modelación, pueden encontrarse en Helbing (2012), Wilensky y Rand (2015) y Rand (2017).

Campos prometedores

En la actualidad es muy fácil capturar y guardar datos, pero es difícil hacer inferencia causal cuando alcanzan cantidades astronómicas, es decir, cuando se convierten en *big data*, sobre todo por limitaciones de hardware. Aún así, dadas las dificultades para conseguir datos experimentales —datos de ensayos aleatorizados— y la hiperabundancia de datos de series de tiempo en la Internet, Google ha estado trabajando desde hace varios años en metodologías de análisis causal bayesiano no experimental. La idea de este análisis estructural de series de tiempo es crear una serie de tiempo contrafactual a partir de datos históricos y compararla con una serie de tiempo factual que presenta un quiebre a partir de una intervención (Brodersen *et al.*, 2015; Brodersen, 2015) y es fácilmente implementable en el software gratuito R (Brodersen y Hauser, 2017). Los usos de esta metodología son más evidentes en el ámbito del marketing y el diseño de páginas web (Tang, Agarwal, O'Brien y Meyer, 2010), pero podrían tener usos en las políticas públicas de la vida cotidiana.

Para poner un ejemplo, se puede suponer que el alcalde desea averiguar qué impacto tendría en la cantidad de riñas callejeras el cierre de una estación de metro cercana a un estadio de fútbol en los días y horarios de partido. Al mismo tiempo se desea no perjudicar las finanzas del metro, que, en el ejemplo, es una empresa pública. La medida del cierre de

la estación se implementó en febrero. A través de un estudio de series de tiempo interrumpidas se puede establecer aproximadamente qué habría pasado si no se hubiera cerrado la estación, haciendo uso de los datos históricos de riñas callejeras de la policía, y contando la cantidad de pasajes que históricamente se han vendido en dicha estación, y la estadísticas de cuántas personas salen de la estación por hora al día. Estadísticamente hablando, este tipo de escenarios son difíciles de modelar, pero haciendo uso de metodologías como las planteadas por el equipo de Google y otras similares es posible hacer una identificación causal aceptable. Se podría estimar, por ejemplo, que gracias a la intervención, entre febrero y diciembre se evitaron 400 riñas y 20 muertes, y se perdieron 500 mil dólares.

Un campo aún más novedoso e inexplorado es el del aprendizaje automatizado (*machine learning*, en inglés). El aprendizaje automatizado, subclase de la inteligencia artificial (IA), refiere a un conjunto de técnicas estadísticas que toman una función matemática y la entrenan reiterativamente hasta que se vuelve experta en hacer predicciones, clasificar datos, o encontrar patrones difíciles de ver para los humanos. Un tipo especial de aprendizaje automatizado, basado en las redes neuronales, es muy usado por ejemplo para predecir los movimientos de corto plazo del mercado accionario y para derrotar a los humanos en diversidad de tareas, tales como en el juego chino *Go* o la conducción de automóviles —los carros que se conducen solos en realidad son conducidos por redes neuronales conectadas entre ellas—. En cuanto a la evaluación de políticas públicas, en el estado del arte de los últimos años se ha visto un avance en dos frentes: la predicción y la selección de covariables.

Una diferencia central entre las técnicas econométricas y las técnicas de aprendizaje automatizado es que en la econometría se busca estimar efectos causales, mientras que en el aprendizaje automatizado no se intentan identificar causalidades, sino sólo hacer predicciones con un alto nivel de precisión. Es por eso que la evaluación de impacto se hace con econometría —“¿fue nuestra política la que *causó* la mejora en los índices de salud?”—, mientras que las tareas puramente predictivas se hacen con inteligencia artificial —“¿cuál será la tasa de desempleo en el próximo trimestre?”—. Sin embargo, esto no significa que el aprendizaje automatizado no tenga utilidad para la evaluación de impacto, toda vez que puede ayudar a evaluar la teoría del cambio a través de predicciones que permitan saber si la teoría funcionará o no. Por ejemplo, puede que en una teoría del cambio una supuesta clave sea que los precios del café se mantendrán estables en un periodo de cuatro meses a partir de la implementación de una política dada. Para evaluar ese supuesto se podría utilizar una red neuronal, por ejemplo, y revisar si las predicciones de la red muestran estabilidad de precios o no. En caso de que la red prediga que los precios estarán estables, ese supuesto se puede aceptar con relativa tranquilidad —que no certeza—; en caso contrario, la teoría del cambio y hasta la intervención misma tendrán que modificarse.

Estas técnicas también se pueden usar para ayudar en la identificación causal. Si bien sus resultados son sólo correlacionales, no causales, con el aprendizaje automatizado se pueden generar hipótesis correlacionales que luego pueden ser testeadas con econometría. Por ejemplo, en una industria dada, si un modelo de predicción de desempleo que incluye

el salario mínimo tiene el mismo poder predictivo que uno que no lo incluye como variable explicativa, el investigador podría entrar a hipotetizar que el salario mínimo no tiene impacto en el nivel de empleo de esa industria.

También se puede usar inteligencia artificial para construir indicadores o puntajes de propensión, con los cuales luego se pueda hacer un estudio econométrico observacional —un emparejamiento estadístico—, o incluso para seleccionar submuestras de interés dentro de la población y hacer experimentos aleatorizados dentro de cada una de ellas (Muralidharan, Cardin, Phillips y Najmi, 2017). Se podría detectar con IA el subconjunto de jóvenes en riesgo de desertar de la universidad, y a través de un experimento aleatorizado entre ellos, evaluar el impacto de un programa de subsidios alimentarios en la tasa de deserción.

En síntesis, en el ámbito de la evaluación de políticas públicas, la inteligencia artificial puede servir para generar hipótesis, para evaluar componentes individuales de la teoría del cambio a través de predicciones, y para facilitar el proceso de construcción de experimentos o cuasi-experimentos. Eso sin contar con que la inteligencia artificial tiene también potencial para aprovechar variables proxy muy creativas. Se puede estimar de manera económica y veloz el PIB de Corea del Norte, por ejemplo, a través de imágenes satelitales procesadas con algoritmos de aprendizaje automatizado, enseñándole al algoritmo a diferenciar entre una casa y una fábrica o a estimar el consumo de los hogares a través de la cantidad de luces encendidas en la noche en el mapa (Mullainathan, 2016; Donaldson y Storeygard, 2016 ; Jean *et al.*, 2016). El aprendizaje automatizado se mezcla, por supuesto, con el *big data*, en aplicaciones como la georreferenciación detallada de la pobreza a través de *Google Street View* (Glaeser, Duke, Luca y Naik, 2015) o usando metadatos de líneas telefónicas celulares (Blumenstock, Cadamuro y On, 2015), y también haciendo *crowdsourcing*, es decir, recopilando datos directamente de la ciudadanía en tiempo real —ya hay aplicaciones que permiten a los ciudadanos reportar obras públicas retrasadas o mobiliario urbano deteriorado a través de sus celulares—. Las oportunidades son muchas, y no sería realista pretender cubrirlas todas en el corto espacio de un artículo.

Conclusiones

Puede arguirse que es fácil perderse en medio de este mar de metodologías, pero en realidad todas ellas están conectadas por el mismo propósito, la mejora continua de la calidad de vida de las personas a través de la mejora continua de las políticas públicas. Esta mejora no necesariamente se encuentra articulada en los planes de desarrollo como son concebidos en la actualidad o en la planificación económica a gran escala a la manera de la época de la Guerra Fría (Friedmann, 1987). De hecho, la tendencia actual en los evaluadores de políticas públicas es a responder preguntas pequeñas, prácticas, poco ambiciosas y que puedan llevar a tomar decisiones específicas con altos niveles de confiabilidad (Duflo y Banerjee, 2012). Es probable, aún así, que con los años la comunidad académica vuelva a adoptar una postura de planificación centralizada más abarcadora, particularmente a medida que se vaya incrementando el corpus de evidencia de políticas públicas alrededor del mundo y se refinen las generalizaciones teóricas.

Todas las metodologías en conjunto son una caja de herramientas, y usualmente no son excluyentes unas con otras, sino complementarias, porque cada una fue creada para resolver complicaciones técnicas específicas. Muchas veces, incluso, es posible hacer evaluaciones muy útiles con poquísimos recursos, usando sólo métodos cualitativos y un poco de sentido común. Es posible y deseable pensar los problemas sociales desde una perspectiva científica, sea cualitativa, cuantitativa o computacional, y a proponer para ellos soluciones responsables. Es cierto que las decisiones finales las toman los representantes políticos, pero los insumos que puedan aportar los ciudadanos pueden hacer la diferencia.

Referencias

- Basu, K. (2013). *The Method of Randomization and the Role of Reasoned Intuition*. The World Bank Group. Recuperado de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2367082
- Bates, M. y Glennerster, R. (2017). The Generalizability Puzzle. *Stanford Social Innovation Review*, 15(3). https://ssir.org/articles/entry/the_generalizability_puzzle
- Berry, D. (2006). Bayesian Clinical Trials. *Nature Reviews. Drug Discovery*, 5, 27-36. <https://www.stat.ubc.ca/~rollin/teach/bar/Berry2006NatureReviewsDD.pdf>
- Blumenstock, J., Cadamuro, G. y On, R. (2015). Predicting Poverty and Wealth from Mobile Phone Metadata. *Science*, 350(6264), 1073-1076. <http://sampleu.ec.unipi.it/wp-content/uploads/2015/12/Science-2015-Blumenstock-1073-6.pdf>
- Brodersen, K. (23 de septiembre de 2015). Causal Attribution in an Era of Big Time-Series Data. [Mensaje en un blog]. Recuperado de <http://www.unofficialgoogledatascience.com/2015/09/causal-attribution-in-era-of-big-time.html>
- Brodersen, K., Galluser, F., Koehler, J., Remy, N. y Scott, S. (2015). Inferring Causal Impact Using Bayesian Structural Time-Series Models. *The Annals of Applied Statistics*, 9(1), 247-274. <https://arxiv.org/pdf/1506.00356.pdf>
- Brodersen, K. y Hauser, A. (2017). *An R Package for Causal Inference Using Bayesian Structural Time-Series Models*. CRAN. Recuperado de <https://cran.r-project.org/web/packages/CausalImpact/vignettes/CausalImpact.html>
- Cartwright, N. (2014). *Evidence: For Policy And Wheresoever Rigor is a Must*. Londres: The London School of Economics and Political Science.
- Cartwright, N. y Munro, E. (2010). The Limitations of Randomized Controlled Trials in Predicting Effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260-266. <https://doi.org/10.1111/j.1365-2753.2010.01382.x>
- Cencini, A. (2005). *Macroeconomic Foundations of Macroeconomics*. Nueva York: Routledge.
- Donaldson, D. y Storeygard, A. (2016). The View From Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4), 171-198. http://dave-donaldson.com/wp-content/uploads/2016/10/Donaldson_Storeygard_JEP.pdf
- Duflo, E. y Banerjee, A. (2012). *Repensar la pobreza: un giro radical en la lucha contra la desigualdad global*. Madrid: Taurus.

- Duflo, E. y Banerjee, A. (eds.). (2017). *Handbook of Field Experiments, Volume 1*. Amsterdam: Elsevier.
- Fienberg, S. (2011). Subjective Bayesian Models and Methods in Public Policy and Government Settings. *Statistical Science*, 26(2), 212-226. <https://doi.org/10.1214/10-STS331>
- Fisher, R. (1937). *The Design of Experiments*. Edimburgo: Oliver and Boyd.
- Forrester, J. (1969). *Urban Dynamics*. Cambridge: M.I.T. Press.
- Forrester, J. (1971). *World Dynamics*. Cambridge: M.I.T. Press.
- Friedmann, J. (1987). *Planning in the Public Domain: From Knowledge to Action*. Princeton: Princeton University Press.
- Fukuyama, F. (1992). *The End of History and The last Man*. Nueva York : The Free Press.
- Ghaffarzadegan, N., Lyneis, J. y Richardson, G. (2011). How Small System Dynamics Models Can Help the Public Policy Process. *System Dynamics Review*, 27(1), 22-44. <https://doi.org/10.1002/sdr.442>
- Giabbanelli, P. y Crutzen, R. (2017). Using Agent-Based Models to Develop Public Policy about Food Behaviours: Future Directions and Recommendations. *Computational and Mathematical Methods in Medicine*, 2017, 1-12. <https://doi.org/10.1155/2017/5742629>
- Ge, F., Wei, Z., Lu, Y., Tian, Y. y Li, L. (2012). Decentralized Coordination Of Autonomous Swarms Inspired by Chaotic Behavior of Ants. *Nonlinear Dynamics*, 70(1), 571-584. <https://doi.org/10.1007/s11071-012-0478-z>
- Gentile, J., Glazner, C. y Koehler, M. (2015). Simulation Models for Public Policy. En B. Furtado, P. Sakowski y M. Tóvolli (Eds.), *Modeling Complex Systems For Public Policies* (pp.73-83). Brasília: Instituto de Pesquisa Econômica Aplicada – IPEA.
- Gertler, P., Martínez, S., Premand, P., Rawlings, L. y Vermeersch, C. (2017). *La evaluación de impacto en la práctica: segunda edición*. Washington: Banco Mundial – Banco Interamericano de Desarrollo.
- Glaeser, E., Duke, S., Luca, M. y Naik, N. (2015). *Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life*. National Bureau of Economic Research. Recuperado de <http://www.nber.org/papers/w21778>
- Glennerster, R. (2017). *When Do Innovation and Evidence Change Lives?* [Mensaje en un blog]. Recuperado de <http://runningres.com/blog/2017/11/9/when-do-innovation-and-evidence-change-lives>
- Haggenson, N. (1990). *System Dynamics Combined with Monte Carlo Simulation*. Proceeds of the 1990 System Dynamics conference. Recuperado de <https://www.systemdynamics.org/assets/conferences/1990/proceed/pdfs/hagen468.pdf>
- Helbing, D. (Ed.). (2012). *Agent-Based Modeling. Social Self-Organization. Understanding Complex Systems*. Berlín: Springer.
- Iacus, S., King, G. y Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1-24. <https://doi.org/10.1093/pan/mpr013>

- Jean, N., Burke, M., Xie, M., Davis, M., Lobell, D. y Ermon, S. (2016). Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science*, 353(6301), 790-794. <https://doi.org/10.1126/science.aaf7894>
- Keane, M. y Wolpin, K. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation: Monte Carlo Evidence. *Review of Economics and Statistics*, 74(4), 648-672.
- Kennedy, W. G., Ihara, E. S., Tompkins, C. J., Inoue, M. y Wolf-Branigin, M. E. (2015). Computational Modeling of Caregiver Stress. *Policy and Complex Systems*, 2(1), 31-44. <https://doi.org/10.18278/jpcs.2.1.5>
- King, G. y Nielsen, R. (2016). *Why Propensity Scores Should Not Be Used for Matching*. Gary King. Recuperado de <https://gking.harvard.edu/files/gking/files/psnot.pdf>
- Lee, D. (2005). An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice. *International Economic Review*, 46(1), 1-34. https://www.jstor.org/stable/3663586?seq=1#page_scan_tab_contents
- Lee, J. y Chu, C. (2012). Bayesian Clinical Trials in Action. *Statistics in Medicine*, 31(25), 2955-2972. <https://doi.org/10.1002/sim.5404>
- Meadows, D. H., Meadows, D. L., Randers, J. y Behrens, W. (1972). *The Limits to Growth*. Nueva York: Universe Books.
- Michaloupoulos, C. (2012). *A Bayesian Reanalysis of Results from the Hard-to-Employ Demonstration and Evaluation Project*. Office of Planning, Research and Evaluation (OPRE) – Administration for Children and Families – U.S. Department of Health and Human Services – Office of the Assistant Secretary for Planning and Evaluation – U.S. Department of Health and Human Services https://www.acf.hhs.gov/sites/default/files/opre/bayesian_reanalysis.pdf
- Mullainathan, S. (2 de abril de 2016). Satellite Images Can Pinpoint Poverty Where Surveys Can't. *The New York Times*. Recuperado de <https://www.nytimes.com/2016/04/03/upshot/satellite-images-can-pinpoint-poverty-where-surveys-cant.html>
- Muralidharan, O., Cardin, N., Phillips, T. y Najmi, A. (31 de enero de 2017). Causality in Machine Learning. [Mensaje en un blog]. Recuperado de <http://www.unofficialgoogledatascience.com/2017/01/causality-in-machine-learning.html>
- Orloff, J. y Bloom, J. (2014). *Comparison of Frequentist and Bayesian Inference*. Massachusetts Institute of Technology. Recuperado de https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading20.pdf
- Pearl, J. (2001). Causal Inference in Statistics: A Gentle Introduction. *Submitted to Computing Science and Statistics, Proceedings of Interface*, 33(1), 1-20. http://ftp.cs.ucla.edu/pub/stat_ser/R289.pdf
- Pearl, J. (2003). Causality: Models, Reasoning and Inference. *Econometric Theory*, 19(4), 675-685.

- Pritchett, L. y Sandfur, J. (2013). *Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix*. Center for Global Development. Recuperado de https://www.cgdev.org/sites/default/files/context-matters-for-size_1.pdf
- Ramsey, F. (1928). A Mathematical Theory of Saving. *The Economic Journal*, 38(152), 543-559.
- Rand, W. (2017). *Introduction to Agent-Based Modeling*. Complexity Explorer – Santa Fe Institute. Recuperado de <https://abm.complexityexplorer.org>
- Ros, J. (2012). La Teoría General de Keynes y la macroeconomía moderna. *Investigación económica*, 71(279), 19-37.
- Salazar, C. (2009). La evaluación y el análisis de políticas públicas. *Ópera*, 9, 23-51.
- Schelling, T. (1978). *Micromotives and Macrobehavior*. Nueva York: W.W Norton.
- Simon, R. (1999). Bayesian Design and Analysis of Active Control Clinical Trials. *Biometrics*, 55(2), 484-487.
- Spiegelhalter, D. y Freedman, L. (1994). Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157(3), 357-416.
- Tang, D., Agarwal, A., O'Brien, D. y Meyer, M. (2010). *Overlapping Experiment Infrastructure: More, Better, Faster Experimentation*. Google. Recuperado de <https://research.google.com/pubs/pub36500.html>
- Todd, P. y Wolpin, K. (2008). Ex-ante Evaluation of Social Programs. *Annales d'Économie et de Statistique*, 91/92, 263-291.
- Wilensky, U. y Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge: The MIT Press.



Cómo citar/ How to cite this item:

Rodríguez, N. S. (2018). Tendencias actuales en la evaluación de políticas públicas. *Ensayos de Economía*, 28(53), 15-35. <https://doi.org/10.15446/ede.v28n53.75382>