

I EVALUACIÓN DE POLÍTICAS PÚBLICAS: INTRODUCCIÓN

J. Ignacio García Pérez
Marisa Hidalgo Hidalgo
Universidad Pablo de Olavide (Sevilla)

Las crecientes limitaciones en la disponibilidad de recursos públicos, así como el cuestionamiento del gasto público con fines sociales, han puesto en primer plano el objetivo de distribuir más eficientemente dichos recursos. Pero la mejora de las políticas públicas y de su eficiencia pasa necesariamente por analizar previamente los efectos de dichas políticas. Países anglosajones (EEUU, Canadá ó Reino Unido) y nórdicos (Suecia ó Dinamarca) han iniciado ya este proceso (Fay, 1996; Larsson, 2003, Sianesi, 2004). Sin duda, este asunto es y será objeto de especial atención por parte de los gestores públicos de las distintas administraciones del Estado.

Por otra parte, el análisis empírico está cada vez más extendido en el ámbito de la investigación económica. La creciente disponibilidad de bases de datos microeconómicas ha hecho avanzar considerablemente la literatura empírica, tanto en aspectos metodológicos como propiamente aplicados. El trabajo empírico que realizan los economistas puede ser dividido en dos grandes grupos: análisis descriptivo e inferencia causal. El análisis descriptivo trata de establecer y documentar hechos sobre los mercados y/o los individuos analizados que necesitan, para su total comprensión, del uso del razonamiento teórico y que, una vez interpretados correctamente, ofrecen información muy valiosa sobre el funcionamiento de la economía. Por otro lado, el análisis de inferencia causal trata de determinar los

efectos de una determinada intervención o política pública, de cara a confirmar las predicciones teóricas sobre el comportamiento individual o de los mercados.

Una de las estrategias tradicionales del análisis económico para la evaluación cuantitativa del efecto de una determinada política pública ha sido el llamado “Enfoque Estructural”. En el marco de un modelo teórico de elección individual y dados unos datos que permiten su estimación empírica, se realiza la evaluación de una determinada política o un cambio de ésta en base a la simulación de sus efectos sobre los parámetros de decisión antes estimados. Éste tipo de análisis es muy útil para evaluar los efectos de posibles cambios de los parámetros de las políticas (por ejemplo, la duración o el tamaño de un subsidio).

En los últimos veinte años, el “Enfoque Causal” o de “efectos de un tratamiento” se ha convertido en un complemento sustancial al anterior. Esto se debe a que ofrece una estrategia de identificación de los efectos causales de una política pública a partir, sobre todo, de datos adecuados para contestar a cada pregunta concreta, más que del soporte estricto de un determinado modelo económico. Este enfoque, más concentrado en los datos y menos necesitado de modelos teóricos o econométricos muy sofisticados, ha generado contribuciones notables en la evaluación de políticas tan relevantes como la formación, los subsidios a la contratación, los efectos de cambios fiscales sobre la oferta de trabajo o la inversión, etc.

Así, la evaluación de políticas públicas constituye hoy un campo de indudable crecimiento, donde el interés científico se ha centrado en campos tan diversos como las finanzas públicas, las políticas de empleo y formación o las subvenciones públicas. Existen numerosos ejemplos en la literatura reciente sobre este tipo de evaluaciones. En particular, sobre el mercado de trabajo han aparecido numerosas contribuciones muy importantes para el desarrollo de esta técnica de evaluación (Card y Krueger, 1994; Ham y LaLonde, 1996; Angrist y Krueger, 1999; Abadie, Angrist e Imbens, 2001).

La pregunta habitual en este tipo de evaluaciones es parecida a la siguiente: ¿Mejora la participación en un programa específico los resultados de los individuos que participan en dicho programa en relación con los resultados que se hubieran observado en esos mismos individuos si no hubiesen participado en el mismo programa específico? Sin embargo, la simple comparación entre individuos sometidos al tratamiento (“grupo de tratamiento”) e individuos no sometidos a él (“grupo de control”) está generalmente sesgada a causa de diferencias no aleatorias entre ambos tipos de individuos que introducen un problema de selección endógena. Los métodos de inferencia causal estudian cómo estimar el efecto causal del programa en estas situaciones.

El problema fundamental de la evaluación de políticas es un problema de identificación. Para ilustrar dicho problema, definamos la variable Y como el resultado

de interés (por ejemplo, el salario), que tiene dos valores “potenciales”. En concreto, Y_1 sería el resultado para un individuo sometido al tratamiento; e Y_0 sería el resultado para ese mismo individuo si no se sometiera al tratamiento. Parece natural, en este contexto, definir el efecto causal del tratamiento para el individuo en cuestión como la diferencia entre los dos resultados potenciales, $(Y_1 - Y_0)$. El problema es que no observamos ambos resultados para el mismo individuo, sino únicamente uno de ellos (Y_1 si el individuo se somete efectivamente al tratamiento e Y_0 en caso contrario). Por tanto, no es posible calcular $Y_1 - Y_0$ para un mismo individuo.

Podría pensarse que este problema de identificación quedaría resuelto comparando los resultados observados de la variable de interés Y , entre individuos tratados (aquellos para los que se observa Y_1) y no tratados (aquellos para los que se observa Y_0). Sin embargo, dicha comparación tampoco proporcionaría información sobre el efecto causal del tratamiento, debido al llamado “problema de auto-selección”. Por ejemplo, los individuos que participan en programas de formación laboral, pueden tener, ya antes de participar en el programa, salarios menores que los que no participan en dicho programa de formación. Por tanto, la diferencia de salarios entre individuos tratados y no tratados resultaría no sólo del tratamiento, sino también de las distintas características de ambos tipos de individuos.

Así por tanto, si los individuos que son candidatos a participar en un programa son asignados aleatoriamente al programa (grupo de tratamiento) o no (grupo de control), como es el caso en los experimentos con un buen diseño y una buena implementación, los posibles resultados del programa serán independientes del mecanismo de asignación, lo que nos permitirá identificar el efecto promedio del tratamiento.

Ahora bien, los datos experimentales son costosos, y no siempre se pueden implementar experimentos (por ejemplo, en caso de que el programa sea un derecho universal, como recibir el subsidio de paro), o diseñar correctamente. Por lo tanto, una alternativa a los análisis experimentales es la utilización de métodos no-experimentales.

El punto clave para usar métodos no experimentales para evaluar el impacto de una determinada política económica es cómo construir un “grupo de comparación” con individuos a los que no se aplica dicha política pero que guardan gran similitud con los individuos a los que se aplica dicha política. El resultado obtenido para la variable de interés por los individuos pertenecientes a este grupo de control se utiliza entonces como medida del resultado atribuible a los participantes en ausencia de política.

Para controlar adecuadamente las diferencias entre el grupo de control y el de tratamiento anteriores a la implementación de la política, se utilizan normalmen-

te modelos estadísticos. Obviamente, el principal problema es que sólo pueden controlarse las diferencias en características observables. Sin embargo, hay un importante número de inobservables (motivación, habilidad, etc.) que pueden afectar a los resultados de la evaluación en la medida en que afecten de manera distinta a los individuos tratados y a los no tratados.

Los sesgos de selección pueden surgir por distintas causas. Como se sugería antes, se pueden deber a decisiones de los propios individuos que solicitan participar en el programa. También se pueden deber a sesgos en las decisiones de selección de los responsables de dichos programas. Para solventar todos estos problemas de selección, existe una amplia variedad de técnicas econométricas desarrolladas en los últimos años. Entre éstas, destacan el uso de variables instrumentales, la modelización del proceso de selección (Heckman, 1979) o el aprovechamiento de la existencia de experimentos naturales, como el usado por Angrist (1990) para evaluar el efecto del servicio militar usando la participación o no en la Guerra de Vietnam. Cabe destacar además la aplicación cada vez más frecuente de métodos consistentes en desarrollar muestras que emparejan participantes y no participantes en una determinada política en base a sus características observables. Este método, conocido como “propensity score”, fue desarrollado por Donald Rubin y ha sido muy utilizado en investigaciones empíricas sobre temas de salud (Rubin, 1973). En los últimos años, esta técnica se ha convertido en la más popular en la literatura económica de evaluación, en especial la dedicada a temas de mercado de trabajo (Dehejia y Wahaba, 1999).

Este libro abunda en cuestiones empíricas y metodológicas relativas a la evaluación de políticas públicas. Así, el siguiente capítulo presenta un análisis de las contribuciones más destacadas tanto en lo que respecta a las distintas metodologías econométricas como a las principales aplicaciones empíricas y los resultados obtenidos en la evaluación microeconómica de políticas públicas. La discusión de este capítulo sobre diseño del muestreo, así como el tipo de información que se requiere para identificar el efecto causal con las distintas metodologías, puede ayudar a establecer qué tipo de datos se necesitan en cada caso. Como Angrist y Krueger (1999) señalan, el futuro del trabajo de investigación aplicado en economía pasa por la “industria incipiente” de generación de bases de datos microeconómicos (basados en registros individuales) que permitan responder a preguntas o predicciones formuladas por los modelos teóricos y que aún no tienen suficiente evidencia empírica que las documente.

El creciente énfasis en datos cuasi-experimentales en el análisis empírico ha hecho ganar importancia al desarrollo de bases de datos para determinados colectivos de la población y/o en determinados períodos para llevar a cabo análisis de inferencia causal en torno a ciertas variables. Con frecuencia, dicho análisis requie-

re muestras muy grandes, en la medida en que sólo pueda explotarse en la estimación una parte de la variación en las variables de interés, que corresponde a una submuestra de la muestra total. El conocimiento exhaustivo de las bases de datos disponibles es esencial para el economista aplicado, al igual que lo es el conocimiento de la metodología o de los modelos teóricos. Concretamente, es imprescindible un amplio conocimiento de la población representada en cada base de datos, del diseño de la muestra (encuesta o registro administrativo), de la tasa de respuesta, de las variables disponibles, del tamaño de la muestra, de la frecuencia de la misma, o , en fin, de los cambios metodológicos de la misma a lo largo del tiempo.

Además, en general, el diseño de las bases de datos disponibles tiene una motivación muy diferente a la del investigador interesado en inferencia causal acerca de una determinada política. En consecuencia, no es extraño que la base de datos utilizada carezca de parte de la información relevante para dicho trabajo de inferencia causal. La falta de dicha información relevante puede deberse al diseño de la base de datos o a la omisión deliberada de dicha información por razones de confidencialidad o de ética profesional en la elaboración de dicha base de datos. Por eso, es cada vez más común que el propio investigador construya los datos necesarios para su investigación mediante el enlace, cuando sea posible, de distintas bases de datos ya existentes, o mediante la realización de encuestas propias que son combinadas con bases de datos ya existentes.

Una fuente de información cada vez más usual son las bases de datos administrativas, esto es, las fuentes de información recogidas por unidades de la administración pública con objetivos propios, es decir, no académicos, pero que pueden ser utilizadas para hacer análisis de inferencia causal. Los ejemplos más usuales de estas bases de datos son los registros de la Seguridad Social, los registros de las oficinas públicas de empleo (pago de prestaciones, cursos de formación, contratos registrados, etc.), los registros de personal de compañías importantes o los registros de las Universidades sobre sus alumnos. La principal ventaja de estas bases de datos es su gran tamaño, lo que permite abordar preguntas muy concretas basándose en submuestras correspondientes a colectivos muy concretos sin riesgo de falta de representatividad en dichos colectivos. Otro aspecto reseñable es que cubren mucha información relevante para analizar decisiones económicas a nivel individual. Por ello, las bases de datos administrativas cumplen una función muy importante en el análisis causal aunque, a veces no proporcionen muestras completamente aleatorias de la población objeto de estudio.

Como ocurre con otras bases de datos procedentes de encuestas, su origen, generalmente ajeno al interés de investigación, repercute a menudo en la falta de disponibilidad de ciertas variables relevantes para el análisis causal (por ejemplo, los registros de la Seguridad Social en España no tienen información sobre el nivel educativo

de los trabajadores). Este problema se soluciona frecuentemente uniendo dichos registros con fuentes de información alternativas, normalmente basadas en encuestas o en otras fuentes administrativas complementarias. Para preservar el problema de la confidencialidad, el enlace de bases de datos se lleva a cabo mediante protocolos que preservan todos los requerimientos legales (ver Duncan y Pearson, 1991). Una estrategia usual, cuando no es posible enlazar bases de datos con información complementaria para individuos comunes, consiste en trabajar de manera separada con ellas y basar la explotación conjunta en la comparación de varios momentos de las distribuciones sujeto de estudio, sin necesidad de que las dos bases de datos estén juntas de cara al análisis multivariante a realizar. A veces, basta disponer de estimaciones separadas de varios momentos de las distribuciones objeto de interés y relacionar económicamente los mismos momentos obtenidos de manera separada (ver, por ejemplo Arellano y Meguir, 1992 o Angrist y Krueger, 1992, 1995).

Al final del Capítulo II, se aborda el análisis Coste-Beneficio de las políticas públicas. Aunque este tipo de análisis no se relaciona estrictamente con la medición de los efectos, permite que las autoridades responsables midan la eficiencia de los programas comparando intervenciones alternativas en función del coste de producir un resultado determinado. De hecho, esta evaluación puede alterar enormemente la cuantificación del impacto y por tanto las implicaciones de la política. Por ello, en el análisis de evaluación de políticas se debería incluir siempre el análisis coste-beneficio (un estudio más completo del análisis coste-beneficio y eficacia en función de los costos se encuentra en el *Handbook on Economic Analysis of Investment Operations*, Banco Mundial, 1996). Con el análisis coste-beneficio se intenta medir, en términos monetarios, la eficiencia económica de los costes de un programa frente a sus beneficios. Sin embargo, para muchos proyectos, especialmente en los sectores sociales, no es posible medir todos los beneficios en términos monetarios.

Además de las principales técnicas y metodologías econométricas para realizar inferencia causal y evaluación de políticas públicas, el capítulo II acompaña cada técnica o metodología de ilustraciones basadas en las contribuciones empíricas más relevantes, para EE.UU. y para algunos países europeos. La discusión más extensa se concentra en los métodos observacionales (matching o emparejamiento, diferencias en diferencias y variables instrumentales), en cuanto que son los de mayor aplicabilidad a la mayor parte de las bases de datos disponibles.

En el Capítulo III se entra de lleno en la descripción de los problemas concretos determinados por los datos en el diseño e implementación de los estudios de evaluación de políticas públicas. Para ello, se describe exhaustivamente el plan de análisis de evaluación de un programa público, poniendo especial énfasis en los problemas de implementación y diseño de las bases de datos necesarias para dicha evaluación.

El Capítulo IV y los siguientes presentan algunos de los trabajos de investigación presentados en las dos jornadas de evaluación de políticas públicas organizadas en el marco del proyecto de investigación desarrollado por el Departamento de Economía de la Universidad Pablo de Olavide y del cual surge el presente libro. Se trata de estudios empíricos donde se evalúan distintas políticas en mercados de trabajo concretos, básicamente en EE.UU., España y algún otro país europeo.

La utilización de los datos y los métodos de recogida en las evaluaciones de políticas en EE.UU. ha experimentado fuertes cambios en los últimos años. El Capítulo IV trata de dar respuesta a las siguientes tres grandes preguntas relacionadas con la recogida de datos en evaluaciones de políticas: ¿Cuáles son las principales ventajas/desventajas de escoger entre las diferentes fuentes de datos? ¿Cuál es la evidencia en EE.UU. respecto a la exactitud de los datos sobre empleo y salarios de diferentes fuentes de datos y cuáles son las implicaciones para la evaluación de políticas? ¿Cuáles son las últimas tendencias en recogida de datos en EE.UU. y hasta qué punto suponen una mejora en la evaluación de políticas? En este capítulo se mostrará que el grado en que nuevos datos se traducen en mejoras en las evaluaciones dependerá de cómo los investigadores continúan desarrollando técnicas de análisis para explotar de forma efectiva estos datos nuevos.

Los capítulos V y VI analizan políticas activas de empleo. El Capítulo V presenta el diseño no-experimental de la evaluación de la efectividad de cuatro políticas activas de ocupación que se ofrecieron en Rumanía en 1999. Estas políticas son: (1) un programa de formación, (2) un programa de apoyo a la auto-ocupación, (3) un programa de creación de trabajo público, y (4) un programa de intermediación laboral. En este capítulo se emplearon datos obtenidos mediante una encuesta que los autores diseñaron exclusivamente para esta evaluación. La metodología utilizada para identificar al grupo de comparación adecuado para cada política consiste en el “matching”. Utilizando dicho método, los autores seleccionan cuatro subgrupos de no-usuarios (uno para cada política) que después del “matching” se parecen mucho a los usuarios en las características laborales y personales antes de participar en el programa. Entre las principales conclusiones está que tres de los cuatro programas (el programa de formación, el de apoyo a la auto-ocupación, y el de intermediación laboral) consiguieron mejorar la situación laboral de sus usuarios. En cambio, dicho estudio encuentra que el programa de creación de trabajo público no mejoró la situación laboral de sus usuarios.

El objetivo del Capítulo VI consiste en el estudio de los servicios de recolocación desarrollados por una de las empresas punteras del sector en España, Creade. Para ello utiliza datos administrativos de dos fuentes diferentes: INEM y Creade. En

este capítulo se plantean diversos métodos de estimación no paramétricos para la evaluación de este tipo de políticas. Cada uno de ellos presenta ventajas e inconvenientes en cuanto a la cantidad y calidad de los emparejamientos. La variable objeto de estudio es el tiempo que un candidato tarda en acceder a un nuevo empleo. El número de días que se reduce el desempleo constituye una variable fundamental para las Administraciones Públicas por lo que representa en la partida de gastos por prestaciones de desempleo. Utilizando este indicador como variable de estudio, las estimaciones sugieren que estos servicios no generan efectos positivos salvo para las mujeres. Esta reacción se debe a lo que el autor denomina el efecto “salario reserva”. Los trabajadores que reciben los servicios de recolocación se hacen más exigentes en el proceso de selección del nuevo puesto de trabajo.

Los capítulos VII y VIII tratan sobre políticas de estabilidad en el empleo. En el Capítulo VII se analizan de qué modo las experiencias de empleo por medio de Empresas de Trabajo Temporal (ETTs) influyen en las carreras laborales de los trabajadores. Para ello, los autores llevan a cabo una evaluación por medio de técnicas de matching del efecto causal que presenta la contratación vía ETTs sobre la probabilidad de alcanzar un contrato indefinido en el futuro. La base de datos utilizada es el Registro de Contratos del Instituto Nacional de Empleo, mediante el cual los autores comparan dos grupos de trabajadores: uno de ellos compuesto por individuos contratados por ETTs (grupo de tratamiento) y el otro compuesto por individuos contratados temporalmente de forma directa (grupo de control). Los resultados obtenidos muestran que los trabajadores que fueron cedidos a empresas clientes sufren una probabilidad significativamente inferior (entre un 15 y un 27%) de alcanzar un contrato indefinido que los contratados temporales directos.

En el Capítulo VIII se evalúa el impacto de las políticas regionales que subvencionan los nuevos contratos indefinidos tomando una muestra de trabajadores con contratos temporales y desempleados para la economía española. En concreto, se estudia la probabilidad de entrada a un contrato indefinido tanto desde el desempleo como desde un contrato temporal a partir de una nueva base de datos procedente de los registros de la Seguridad Social en España denominada la “Muestra Continua de Vidas Laborales”, que se utiliza por primera vez para realizar ejercicios de evaluación de políticas públicas en el mercado de trabajo español. En cuanto a la metodología empleada, los autores parten de los modelos de duración para identificar el efecto de la política sobre la probabilidad de transitar a un contrato indefinido. Por otra parte, dado que su base de datos ofrece información longitudinal y su variable de política tiene variabilidad individual, regional y temporal aplicarán el método de triples diferencias para identificar el efecto medio de la política. Sus principales resultados son que, en media, esta política tiene efectos positivos,

pero pequeños, en la probabilidad de transición a un contrato permanente, tanto desde el desempleo como desde un contrato temporal. No obstante, la incidencia de esta política difiere notablemente entre hombres y mujeres y en menor medida entre los grupos de edad. Es mayor para las mujeres mientras que para los hombres esta política no tiene impacto positivo relevante.

Los capítulos IX y X realizan evaluaciones de políticas educativas. En el Capítulo IX se obtienen estimaciones de rendimientos educativos. Asimismo se contrasta su heterogeneidad, bajo distintos supuestos, y usando una muestra de corte transversal de una encuesta muy interesante de datos españoles (*Encuesta de Conciencia y Biografía de Clase*, 1991). Los métodos utilizados son de estimación semiparamétrica. El principal resultado que presenta este capítulo es que puede haber grupos de la población para los cuales los efectos marginales del tratamiento son negativos. En este caso, estos individuos no se beneficiarán, en términos de salarios, de un aumento en sus niveles educativos. Esto es algo que se debe tener en cuenta a la hora de evaluar políticas que obligan a los individuos a adquirir un determinado nivel de educación. Una aplicación al análisis de la reforma educativa, que tuvo lugar en España en 1970, sugiere que aquellos individuos que estudiaron antes de la reforma, y que tienen una mayor probabilidad *a priori* de seguir estudiando a partir del nivel educativo bajo (a partir de los 14 años), se hubieran beneficiado, en términos de salarios obtenidos, de esta reforma.

El Capítulo X realiza una estimación de los efectos a corto, medio y largo plazo de diferentes tipos de Formación Subvencionada por el Gobierno sobre el empleo, desempleo e ingresos de los participantes en Alemania del Este y el Oeste. Estos programas difieren considerablemente tanto en la forma y la intensidad de la inversión en capital humano así como en su respectiva duración. La base de datos empleada en este estudio combina tres fuentes diferentes: una muestra de empleados, el registro de pago de subsidios y los datos de participantes en formación. Estos datos administrativos que usan los autores son particularmente ricos lo que les permite corregir por la selección por métodos de emparejamiento que tienen en cuenta múltiples tratamientos y que miden interesantes variables de resultado en base mensual para un período de ocho años después de la entrada en el programa. Para una muestra de desempleados del este y oeste de Alemania que entraron en programas de formación en 1993 y 1994 los autores encuentran que todos los tipos de formación que evalúan tienen efectos negativos a corto plazo sobre el empleo cuya magnitud y duración están directamente relacionadas a la duración del programa tanto para el Este como el Oeste de Alemania. Sin embargo, a largo plazo todos los tipos de programa parecen incrementar las tasas de empleo en aproximadamente 10 puntos porcentuales. Esto se mantiene no sólo para los participan-

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

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