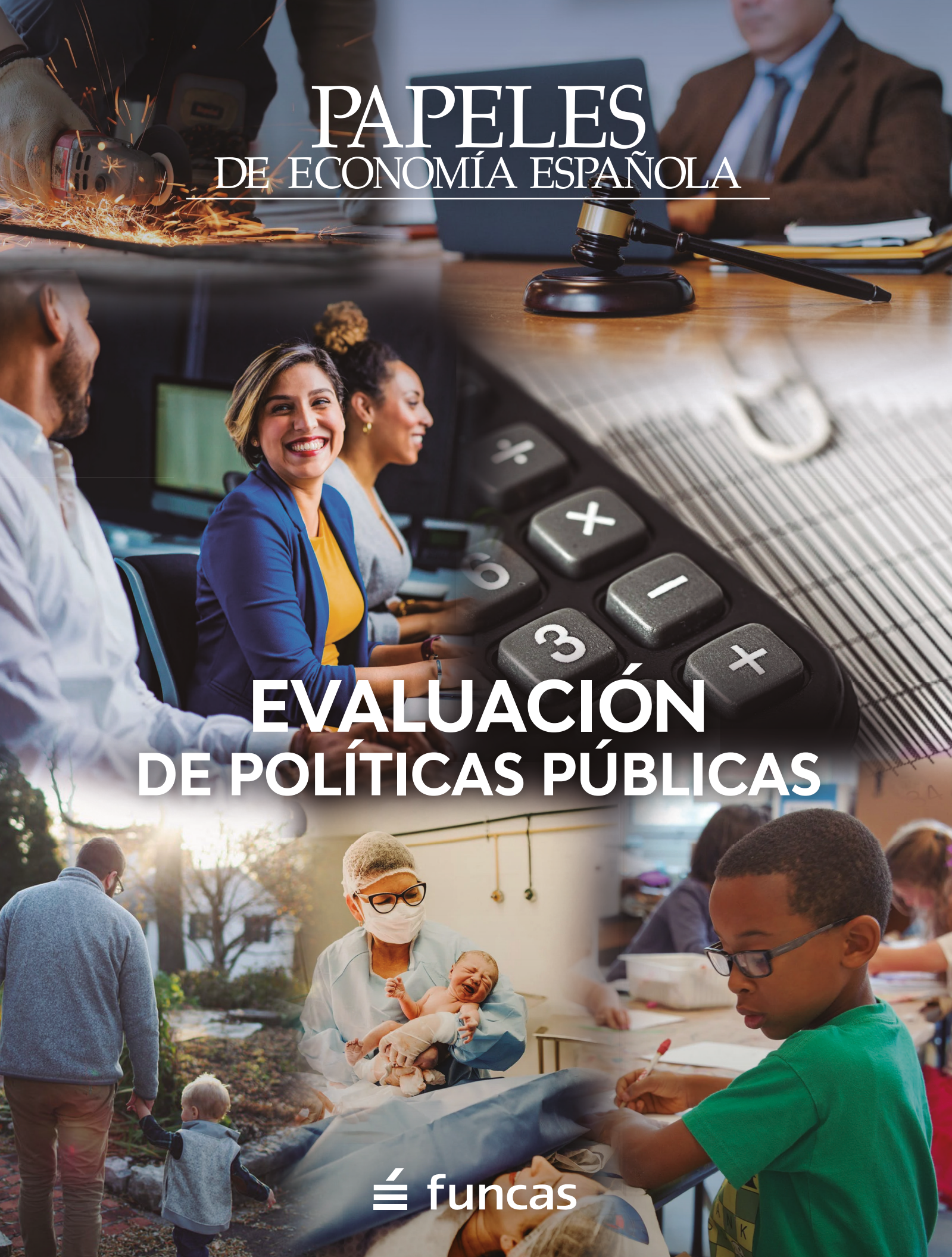




PAPELES DE ECONOMÍA ESPAÑOLA

EVALUACIÓN DE POLÍTICAS PÚBLICAS

PAPELES DE ECONOMÍA ESPAÑOLA es una revista que edita trimestralmente Funcas. La revista pretende ofrecer una información rigurosa y responsable de los problemas económicos españoles. Cinco son los criterios a los que PAPELES DE ECONOMÍA ESPAÑOLA desea ajustar su contenido.

- La plena libertad intelectual de quienes colaboran en sus páginas.
- La responsabilidad de las opiniones emitidas y el respaldo riguroso de las afirmaciones realizadas mediante datos y pruebas que avalen los planteamientos efectuados y las soluciones propuestas.
- El compromiso de las opiniones con los problemas planteados. La economía española se encuentra ante uno de los mayores retos de su historia, y dar opiniones comprometidas en estos momentos es obligación ineludible de los profesionales más cualificados.
- La búsqueda y la defensa de los intereses generales en los problemas planteados.
- La colaboración crítica de los lectores, para que nuestros Papeles se abran al diálogo y la polémica, si fuera preciso, para buscar soluciones a problemas de la economía española.

PAPELES DE ECONOMÍA ESPAÑOLA se ofrece como un medio de expresión a cuantos se interesen por nuestros temas económicos, cumpliendo con el único requisito de la previa aceptación de las colaboraciones remitidas por el Consejo de Redacción que, sin embargo, respetuoso con la libertad intelectual no modificará las ideas, opiniones y juicios expresados por los autores ni tampoco se solidarizará con ellas.

PAPELES DE ECONOMÍA ESPAÑOLA está abierta a la colaboración de todos los investigadores que deseen aportar sus trabajos sobre temas referentes a la economía española. Los envíos de originales deberán hacerse al Departamento de Edición y Documentación de Funcas, Caballero de Gracia, 28. 28013 Madrid.

PAPELES
DE ECONOMÍA ESPAÑOLA

172

2022

ISSN: 0210-9107



PATRONATO

ISIDRO FAINÉ CASAS (<i>Presidente</i>)	MANUEL AZUAGA MORENO
JOSÉ MARÍA MÉNDEZ ÁLVAREZ-CEDRÓN (<i>Vicepresidente</i>)	CARLOS EGEA KRAUEL
FERNANDO CONLLEDO LANTERO (<i>Secretario</i>)	MIGUEL ÁNGEL ESCOTET ÁLVAREZ
	AMADO FRANCO LAHOZ
	PEDRO ANTONIO MERINO GARCÍA
	ANTONIO PULIDO GUTIÉRREZ
	VÍCTORIO VALLE SÁNCHEZ
	GREGORIO VILLALABEITIA GALARRAGA

PAPELES DE ECONOMÍA ESPAÑOLA

DIRECTOR

Eduardo Bandrés Moliné

CONSEJO DE REDACCIÓN

CARLOS OCAÑA PÉREZ DE TUDELA (<i>Director</i>)	ELISA CHULIÁ RODRIGO
SANTIAGO CARBÓ VALVERDE	JUAN JOSÉ GANUZA
	RAYMOND TORRES

COORDINADORA DE EDICIÓN Y DOCUMENTACIÓN

Myriam González Martínez

PORTADA

©Advantia, Comunicación Gráfica, S.A.

EDITA

Funcas
Caballero de Gracia, 28. 28013 Madrid

IMPRIME

Advantia Comunicación Gráfica, S.A.

Depósito legal:	M. 402-1980
ISSN:	0210-9107
Precio del número impreso:	20 €
Versión digital:	Gratuita
Periodicidad:	Trimestral
Materia:	Economía Pública
Disponible en formato digital:	www.funcas.es



© FUNCAS. Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta publicación, así como la edición de su contenido por medio de cualquier proceso reprográfico o fónico, electrónico o mecánico, especialmente imprenta, fotocopia, microfilm, *offset* o mimeógrafo, sin la previa autorización escrita del editor.

«Evaluación de políticas públicas»

Coordinado por Francisco Pedraja Chaparro

SUMARIO

INTRODUCCIÓN EDITORIAL

Evaluación de políticas públicas

V

COLABORACIONES

I. ASPECTOS GENERALES

Similitudes y diferencias en la evaluación de políticas públicas en ocho países: España en perspectiva comparada:	2	José María Casado y Eloísa del Pino
Métodos de evaluación de políticas públicas:	18	Joaquín Artés y Beatriz Rodríguez-Sánchez
Consecuencias no intencionadas de las medidas de política pública: ejemplos de sanidad y educación:	30	Jennifer Graves y Eugenio Zucchelli
Desarrollos recientes en eficiencia y productividad para su aplicación en el sector público:	43	Daniel Santín González

II. IMPUESTOS Y POLÍTICA FISCAL

¿Más o menos impuestos? Renta, confianza en las instituciones y preferencias impositivas:	58	Carlos Sanz
Evaluación del impacto de un incremento del IVA reducido sobre el gasto de los hogares en España:	69	Ignacio Moral-Arce
Los planes de política fiscal: evaluación <i>ex ante</i> y <i>ex post</i> :	85	Iván Kataryniuk Lucía López Muñoz y Javier J. Pérez

III. GASTO PÚBLICO

La evaluación de las políticas públicas: la administración de justicia:	100	Juan S. Mora-Sanguinetti
Evaluación de políticas. Salud y sanidad:	110	Beatriz González López-Valcárcel

La evaluación en el contexto educativo:	126	<i>José Manuel Cordero Ferrera Francisco Pedraja Chaparro y Rosa Simancas Rodríguez</i>
La evaluación del impacto de las políticas públicas: el caso de las políticas del mercado de trabajo:	139	<i>José M. Arranz y Carlos García-Serrano</i>
El ingreso mínimo vital: adecuación y cobertura:	155	<i>Luis Ayala Cañón Antonio Jurado Málaga y Jesús Pérez Mayo</i>

EVALUACIÓN DE POLÍTICAS PÚBLICAS

INTRODUCCIÓN EDITORIAL

LA justificación de la evaluación de las políticas públicas puede hacerse, al menos, desde tres dimensiones. En primer lugar, la rendición de cuentas a los ciudadanos en el uso de los fondos públicos por parte de sus gobernantes es una pieza fundamental en toda democracia. En segundo lugar, los principios de eficacia y eficiencia en el uso de los recursos públicos aparecen recogidos en nuestras normas de mayor rango (Constitución Española, Ley Orgánica de Estabilidad Presupuestaria y Sostenibilidad Financiera y Ley General Presupuestaria, por poner tres ejemplos relevantes). Por último, desde una perspectiva más pragmática, el diseño correcto de la política pública, esto es, conocer si alcanza sus objetivos y, en caso contrario, proponer su modificación o supresión exige su evaluación de manera que los programas públicos se basen en resultados y no en buenas intenciones o meras intuiciones.

El desarrollo de la evaluación de las políticas públicas, como le ha sucedido a la propia evolución de la Economía Pública en las últimas décadas como una disciplina cada vez más empírica (Martínez-Vázquez, 2018), se ha visto favorecido por dos elementos claramente relacionados. Por una parte, la disponibilidad de microdatos administrativos de carácter longitudinal. Por otra, el desarrollo de nuevas técnicas, en especial, el uso de experimentos aleatorios (*Randomized Controlled Trials* o *RCTs*), recientemente refrendadas con la concesión de varios premios nobel de Economía (Banerjee, Duflo y Kremer en 2019 y Angrist, Card e Imbens en 2021).

Si descendemos a nuestro país, desde una perspectiva institucional en términos comparados, España alcanza un nivel medio-bajo en la evaluación de políticas públicas (Stockmann, Meyer y Taube, 2020). Lo cierto es que nos falta una verdadera cultura de la evaluación como sucede en los países con una larga tradición en evaluación (Reino Unido y Estados Unidos).

Las bases de datos, en su mayoría, no están diseñadas para evaluar sino con objetivos más tradicionales de control del gasto y, en general, no se encuentran a disposición de los investigadores. En ese sentido, una interesante iniciativa es la propuesta de la Autoridad Indepen-

EL DISEÑO CORRECTO DE LA POLÍTICA PÚBLICA EXIGE SU EVALUACIÓN DE MANERA QUE LOS PROGRAMAS PÚBLICOS SE BASEN EN RESULTADOS Y NO EN BUENAS INTENCIONES O MERAS INTUICIONES

EL DESARROLLO DE LA EVALUACIÓN DE LAS POLÍTICAS PÚBLICAS SE HA VISTO FAVORECIDO POR DOS ELEMENTOS: LA DISPONIBILIDAD DE MICRODATOS ADMINISTRATIVOS DE CARÁCTER LONGITUDINAL Y EL DESARROLLO DE NUEVAS TÉCNICAS, EN ESPECIAL, EL USO DE EXPERIMENTOS ALEATORIOS

**NOS FALTA UNA
VERDADERA CULTURA
DE LA EVALUACIÓN
COMO SUCEDER EN
LOS PAÍSES CON UNA
LARGA TRADICIÓN EN
EVALUACIÓN (REINO
UNIDO Y ESTADOS
UNIDOS)**

diente de Responsabilidad Fiscal (AIReF, 2020) de llevar a cabo una estrategia nacional para la disponibilidad y el uso de los datos que contienen los registros administrativos.

En cuanto a las técnicas más novedosas de evaluación que ponen el énfasis en el establecimiento de causalidad entre los programas públicos y sus resultados e impactos, no existen especialistas en la administración pública y se encuentran al alcance solo de unos pocos académicos universitarios que las utilizan en sus investigaciones.

La situación sin embargo parece que está cambiando. Un paso relevante en la buena dirección ha sido la puesta en marcha de la División de Evaluación del Gasto Público en la AIReF en 2021, consolidando con ello la actividad que esta institución venía realizando con los llamados *Spending Reviews* desde 2017. Por otra parte, el Plan de Recuperación, Transformación y Resiliencia constituye una buena oportunidad para profundizar en la práctica evaluadora a juzgar por la forma en la que deben justificarse la percepción de los fondos asociados a ese plan.

Un número de *Papeles de Economía Española* dedicado a la «Evaluación de las Políticas Públicas» parece, en consecuencia, plenamente justificado. El que presentamos consta de doce trabajos agrupados en tres bloques; cuatro forman la parte introductoria, tres se dedican a imposición y política fiscal, y cinco dan el contenido a las políticas de gasto.

La sección introductoria comienza con un análisis comparado de la evaluación de las políticas públicas en España con ocho países caracterizados por su tradición evaluadora por parte de **José María Casado** y **Eloísa del Pino**. Los autores constatan el avance que se ha producido en nuestro país en la institucionalización de la evaluación de sus políticas públicas, así como una serie de deficiencias y retos importantes que agrupan en diferentes aspectos. En relación al marco organizativo y regulatorio de la evaluación, no existe en nuestro país una ley de evaluación de políticas, aunque sí normas que la mencionan de manera más o menos directa como sucede en la mayoría de los países. Un paso importante en esta dirección, después de la escasa exitosa experiencia de la Agencia Estatal de Evaluación de la Calidad de los Servicios y de las Políticas Públicas (AEVAL), es la puesta en marcha de la División de Evaluación del Gasto Público en la Autoridad Independiente de Responsabilidad Fiscal (AIReF) en 2021 que ha formalizado la práctica de la evaluación en esa institución independiente

**UN PASO RELEVANTE
HA SIDO LA
PUESTA EN MARCHA
DE LA DIVISIÓN DE
EVALUACIÓN DEL
GASTO PÚBLICO EN LA
AIReF EN 2021**

que llevaba desarrollándose desde 2017 y que se consolidará aún más con el papel que tendrá la institución como organismo evaluador en el Plan de Recuperación, Transformación y Resiliencia (PRTR) del Gobierno central (2021). En cuanto a la capacidad evaluadora, se pone de manifiesto la relevancia de disponer de datos administrativos en la experiencia comparada y el reto que supone para nuestro país la puesta en marcha de una estrategia nacional con ese objetivo, así como disponer de efectivos en la administración con conocimientos específicos en técnicas de evaluación. Por último, una dimensión relevante es el uso de las evaluaciones en el que queda mucho camino por recorrer si nos comparamos con países en los que la publicación de los resultados de las evaluaciones es obligatoria y se hace un seguimiento de tales resultados. En este sentido, el PRTR incluye la creación de una unidad en el Ministerio de Hacienda para el seguimiento de las propuestas de la AIREF y su incorporación al ciclo presupuestario. También ayudará a la difusión el que las evaluaciones de la AIREF sean llevadas a las conferencias sectoriales donde se reúnen los ministerios y consejerías autonómicas correspondientes.

EL PRTR INCLUYE LA CREACIÓN DE UNA UNIDAD EN EL MINISTERIO DE HACIENDA PARA EL SEGUIMIENTO DE LAS PROPUESTAS DE LA AIREF Y SU INCORPORACIÓN AL CICLO PRESUPUESTARIO

Dado el extraordinario desarrollo producido en los últimos años en las técnicas de evaluación de impacto, era obligado dedicar un trabajo a ofrecer una panorámica de las mismas, tarea que llevan a cabo **Joaquín Artés** y **Beatriz Rodríguez-Sánchez**. A pesar de la complejidad del tema, el resultado es óptimo debido a la claridad de su contenido y al apoyo de ejemplos con el que ilustran las diferentes alternativas. Tras plantear el problema fundamental de identificación de relaciones causales en el ámbito de las políticas públicas, se aborda el método de aleatorización que permite asegurar que el grupo tratado y el grupo de control sean estadísticamente idénticos en todo aquello que no es la pertenencia al grupo tratado, lo que permite atribuir las diferencias entre los dos grupos al tratamiento o a la política aplicada. A veces la aleatorización se produce de modo natural (involuntariamente) mediante experimentos naturales. Cuando lo anterior no es posible, como sucede en la mayoría de las evaluaciones de las políticas públicas, hay que recurrir a otras técnicas estadísticas. Las más destacadas son las cuasiexperimentales que permiten identificar los efectos causales incluso cuando existen diferencias no observables entre el grupo tratado y el de control. De ellas se ocupa el trabajo en su parte final (regresión discontinua, diferencias en diferencias y variables instrumentales) destacando sus ventajas e inconvenientes en relación con las características de la política a evaluar y con la información disponible en cada caso.

A VECES LA ALEATORIZACIÓN SE PRODUCE DE MODO NATURAL (INVOLUNTARIAMENTE) MEDIANTE EXPERIMENTOS NATURALES

El artículo de **Jennifer Graves** y **Eugenio Zucchelli** versa sobre un aspecto tan relevante como poco tratado en el ámbito de la evaluación de las políticas públicas, la necesidad de considerar las consecuencias inesperadas e indeseables al evaluar el efecto causal de tales políticas, que los autores ilustran mediante una serie de ejemplos para los casos de sanidad y educación. En políticas sanitarias se destaca la importancia de incluir en la evaluación los efectos sustitución entre bienes, los efectos indirectos en otros sectores y tener en cuenta las interacciones que se producen entre las legislaciones de distintos niveles de gobierno sobre una determinada materia. En políticas educativas se insiste en los incentivos que se producen en las respuestas a determinados objetivos (resultados en los exámenes) con la reducción del tiempo dedicado a desarrollar contenidos o capacidades no evaluadas tanto o más importantes. También se llama la atención sobre los problemas de implementar políticas a gran escala a partir de los resultados obtenidos al aplicar una política en un ámbito reducido, como por ejemplo la reducción del número de alumnos por aula cuyos efectos beneficiosos acaban más que contrarrestándose con una pérdida de calidad del profesorado que ha de contratarse masivamente para aplicar esa política. Como reflexión final, se aboga por una evaluación periódica de las políticas teniendo en cuenta que buena parte de esos efectos se producen pasados bastantes años desde que las políticas públicas se aplican.

Cerramos la parte introductoria con un trabajo dedicado a los desarrollos recientes en eficiencia y productividad en el ámbito del sector público a cargo de **Daniel Santín González**. A este tema ya en 1995 la revista *Papeles de Economía Española* dedicó su número 95 con el título «Sector Público y Eficiencia» también coordinado por el profesor Pedraja. Tras destacar la conexión entre eficiencia y evaluación y la importancia de mejorar aquella en nuestro sector público, dada su delicada situación financiera y la relevancia cuantitativa de los programas de gasto representativos del Estado del bienestar, el artículo expone tres desarrollos metodológicos recientes en los métodos de frontera en los que el autor ha tenido un papel protagonista. En primer lugar, las metafronteras y tecnologías de referencia, de enorme interés para estados descentralizados como sucede en nuestro país. En segundo lugar, la combinación de las herramientas de la inferencia causal con las técnicas de medición de la eficiencia basadas en el uso de fronteras. Por último, la utilización del análisis de fronteras productivas en la cuantificación de ganancias de eficiencia en procesos de fusión de unidades que prestan servicios similares. El artículo se cierra con la

propuesta de crear agencias de evaluación en el ámbito autonómico en España.

La sección de imposición y política fiscal comienza con el trabajo de **Carlos Sanz** sobre preferencias impositivas, renta de los contribuyentes y confianza en las instituciones. Se trata de un valioso análisis descriptivo, complementario de posteriores análisis causales, en el que se obtienen interesantes resultados a partir de una encuesta. Son los hogares con más renta los que expresan una mayor preferencia por subir los impuestos, lo que contradice el marco teórico tradicional en el caso de que se asocien mayores impuestos a más redistribución. El estudio muestra además que otros factores sociodemográficos y políticos correlacionan con las preferencias impositivas de los ciudadanos; en ese sentido, los hombres, las personas de más edad y los votantes de izquierda muestran una mayor preferencia por subir los impuestos. Por último, una variable clave para explicar las preferencias impositivas es la confianza en las instituciones públicas y el sistema político, contribuyendo a apoyar la literatura que identifica a ese factor de la confianza en las instituciones como un elemento crucial para el desarrollo económico y el buen funcionamiento institucional.

El artículo de **Ignacio Moral-Arce** se ocupa del impacto de las subidas de tipos reducidos en el IVA en los años 2010 y 2012. Tras repasar los estudios que analizan el impacto de los cambios en la imposición indirecta general y especial en España, utilizando tanto modelos de simulación como aproximaciones contrafactuales, analiza los efectos de esos cambios utilizando la EPF en el periodo 2007 a 2015 con técnicas de regresión en discontinuidad. Para ello distingue lo que denomina margen extensivo, número de hogares afectados, y margen intensivo o efecto sobre el gasto medio de los hogares, diferenciando el impacto de cada reforma por grupos de bienes. Un resultado interesante es el distinto efecto, por niveles de ingresos del hogar, de las reformas de 2010 y 2012 por categorías de bienes.

Al ámbito de la evaluación de la política fiscal se dedica el trabajo de **Iván Kataryniuk, Lucía López Muñoz y Javier J. Pérez** en el que, desde una perspectiva macroeconómica, se revisan los principales elementos institucionales y técnicos en la evaluación *ex ante* y *ex post* de los planes de la política fiscal. En primer lugar, se repasa el marco institucional europeo y nacional para la elaboración y supervisión de los planes fiscales, incluyendo una referencia particular al papel de los organismos de control en los mismos. En segundo lugar, se exponen

**EL PROGRAMA NGEU
ES UN EJEMPLO
PROMETEDOR
DE DISEÑO
INSTITUCIONAL
QUE INCORPORA
INCENTIVOS Y
ELEMENTOS
DE VALORACIÓN
EX ANTE Y EX POST**

los elementos técnicos básicos necesarios para una adecuada evaluación *ex ante* de los planes. En tercer lugar, se examinan los aspectos más relevantes de acuerdo con la literatura de referencia sobre la evaluación *ex post* de los planes, elemento crucial para dotar de credibilidad a los mismos y pieza esencial de la rendición de cuentas de los gobiernos frente a los ciudadanos, resaltando el caso del programa NGEU como un ejemplo prometedor de diseño institucional que incorpora incentivos y elementos de valoración *ex ante* y *ex post*.

La evaluación de las políticas públicas en la administración de justicia corre a cargo de **Juan S. Mora-Sanguinetti**, quien divide el contenido de su trabajo en cuatro secciones. En la primera, justifica la evaluación del buen funcionamiento del sistema judicial poniendo el énfasis en su relevancia económica en sus dos vertientes privada y pública. En este sentido, se insiste en la relación entre seguridad jurídica e inversión y se llama la atención sobre cómo la acción de la justicia, desde una perspectiva global, determina, en parte, el comportamiento de la productividad. En la segunda sección, tras repasar los indicadores que la literatura utiliza para evaluar la eficacia de la justicia, se decanta por su lentitud como el indicador más adecuado y repasa las fuentes de información existente para un análisis comparado entre países. En la tercera sección, a partir de las fuentes anteriores, ofrece un panorama de la situación de la justicia española y pone de relieve su elevada tasa de congestión general, así como las notables diferencias entre comunidades autónomas. Por último, mediante ejemplos de impactos económicos se analizan algunos problemas de funcionamiento de la justicia en España: marco regulatorio complejo y negativas implicaciones para la inversión de la ineficacia de la jurisdicción civil.

El artículo de **Beatriz González López-Valcárcel** constituye una excelente reflexión sobre datos y métodos de evaluación de políticas sanitarias y de salud, su organización en España y los temas relevantes para una agenda priorizada. La autora señala que, pese al enorme avance en la disponibilidad de datos, la metodología para procesarlos, para evaluar políticas públicas, está relativamente estancada. La modelización ha progresado, pero todavía falta integrar mejor los modelos epidemiológicos y económicos, así como incorporar los efectos sobre la desigualdad. En cuanto a la gestión de la pandemia, pone de manifiesto la falta de una evaluación global de la misma, que actualmente se encuentra en proceso de elaboración. La evaluación de la sanidad debe abarcar tres niveles: macro (políticas de salud y sanitarias, meso (gestión y organización de centros y servicios sanitarios) y micro (gestión

**LA MODELIZACIÓN
HA PROGRESADO,
PERO TODAVÍA FALTA
INTEGRAR MEJOR
LOS MODELOS
EPIDEMIOLÓGICOS
Y ECONÓMICOS, ASÍ
COMO INCORPORAR
LOS EFECTOS SOBRE
LA DESIGUALDAD**

clínica), siendo áreas de especial relevancia para la evaluación la digitalización de la sanidad, la gestión indirecta de la prestación sanitaria y las tecnologías de alto coste unitario, como la medicina personalizada, y finaliza reclamando, entre otras acciones, una evaluación urgente del modelo MUFACE.

El artículo de **José Manuel Cordero Ferrara, Francisco Pedraja Chaparro y Rosa Simancas Rodríguez** ofrece una panorámica de la evaluación en educación. Comienza justificando la importancia de la evaluación en educación y, concretamente, en los resultados educativos ya que explican una parte importante de las tasas de crecimiento de los países. A continuación, se repasan, sin un ánimo exhaustivo, algunos estudios que utilizan las técnicas de evaluación de impacto en la función de producción educativa tanto desde una perspectiva internacional como nacional. Los autores ponen de manifiesto la superioridad de los experimentos aleatorios controlados como método de evaluación cada vez más utilizados especialmente en países con mayor tradición evaluadora. Precisamente esa es la gran asignatura pendiente de España en la evaluación de políticas educativas, la utilización de experimentos aleatorios, aunque existen algunas excepciones impulsadas por diversas instituciones. Por último, a modo de ejemplo, evalúan el Programa 18-25 implementado en Extremadura consistente en la oferta de un incentivo económico a los desempleados con edades comprendidas entre 18 y 25 años para que finalizasen la enseñanza secundaria obligatoria e incrementasen con ello sus posibilidades de incorporación al mundo laboral. Mediante el uso de una regresión en discontinuidad, se demuestra que la posibilidad de obtención del título de la ESO no fue estadísticamente diferente entre el grupo de tratados y el grupo de control. En definitiva, esa política educativa, no resultó efectiva.

La evaluación de las políticas en el mercado de trabajo corre a cargo de **José María Arranz y Carlos García-Serrano**. Su propósito es doble, hacer una revisión de estos trabajos para el caso español y mostrar un ejemplo práctico de evaluación de alguna política realizada en el mercado de trabajo en nuestro país. En cuanto al primer objetivo, ponen de manifiesto la escasez de estudios que aplican estas técnicas apuntando un par de razones. Por una parte, la dificultad de disponer de datos adecuados (longitudinales y representativos); por otra, la simultaneidad de cambios en la normativa laboral que imposibilita la delimitación adecuada entre grupos de tratados y de control. El ejercicio de evaluación se centra en el cambio legislativo producido en 2012 por el que se aumentó la edad mínima para solicitar una

EL PROGRAMA 18-25 IMPLEMENTADO EN EXTREMADURA DEMUESTRA QUE LA POSIBILIDAD DE OBTENCIÓN DEL TÍTULO DE LA ESO NO FUE ESTADÍSTICAMENTE DIFERENTE ENTRE EL GRUPO DE TRATADOS Y EL GRUPO DE CONTROL

EL CAMBIO LEGISLATIVO PRODUCIDO EN 2012 POR EL QUE SE AUMENTÓ LA EDAD MÍNIMA PARA SOLICITAR UNA PRESTACIÓN ASISTENCIAL ILIMITADA TUVO EFECTOS SUSTANCIALES EN EL COMPORTAMIENTO DE LOS TRABAJADORES MAYORES

prestación asistencial ilimitada. Utilizando microdatos del Servicio Público de Empleo Estatal (SEPE) y mediante una técnica de diferencias en diferencias, llegan a la conclusión, análoga a la obtenida en la literatura comparada, de que la política tuvo efectos sustanciales en el comportamiento de los trabajadores mayores. Concretamente, la probabilidad de salir del desempleo hacia un empleo para las personas de 52 a 54 años que dejaron de tener acceso a los subsidios ilimitados después de que se produjo el cambio de política, fue 2,5 veces mayor que la que había antes de ese cambio.

Para finalizar, en el artículo de **Luis Ayala Cañón, Antonio Jurado Málaga y Jesús Pérez Mayo** se lleva a cabo una primera valoración del ingreso mínimo vital (IMV) puesto en marcha en junio de 2020, una importante novedad ya que se trata de la primera renta mínima de ámbito nacional en España. El tema es muy relevante dados los elevados niveles, en términos comparados europeos, de nuestro país en pobreza y desigualdad y las notables diferencias por categorías demográficas y territoriales que existen en España, donde buena parte de las competencias en esta política social se encuentra descentralizada en las CC. AA. Con la escasa información disponible y el breve periodo de desarrollo del IMV, el trabajo utiliza diversas fuentes como los microdatos de la ECV 2020, la encuesta FOESSA 2021 y los registros administrativos. Entre sus principales conclusiones se destaca que, pese a la mejora en relación con los niveles ofrecidos por los esquemas autonómicos, se observa el escaso alcance del programa para la población objetivo que se habría duplicado de haber utilizado el umbral del 40 por 100 de la mediana de la renta equivalente del hogar, el más frecuente en la literatura especializada. De los resultados también se desprende que la principal causa de denegación es el incumplimiento de los umbrales económicos de ingresos y patrimonio, por lo que resulta imprescindible una modificación del modo de determinar los recursos económicos a evaluar. Por último, se identifican problemas considerados de ineficiencia asignativa vertical al existir un significativo porcentaje de unidades que, sin pertenecer a la población objetivo, declaran recibir la prestación, lo que pone de manifiesto un evidente problema de equidad del programa.

BIBLIOGRAFÍA

- MARTÍNEZ-VÁZQUEZ, J. (2018). Perspectives on the Last Quarter Century of Research in Public Economics. *Revista de Economía Aplicada*, 76 (vol. XXVI), pp. 9-33.
- STOCKMANN, R., MEYER, W. y TAUBE, L. (eds.) (2020). *The Institutionalization of Evaluation in Europe*. Cham: Palgrave Macmillan.

COLABORACIONES

I. ASPECTOS GENERALES

Resumen

En este artículo se analiza la situación de la evaluación de políticas públicas en ocho países, entre los cuales se encuentra España. En concreto, se explora la institucionalización de la evaluación en el sistema político-administrativo y en particular su marco normativo y organizativo, la capacidad de evaluación de las administraciones y el uso que se hace de la evaluación por parte de los decisores públicos. Los países se han escogido atendiendo a diversos criterios: larga tradición de evaluación (Reino Unido y Estados Unidos), países con un alto grado de institucionalización de la evaluación en sus sistemas políticos (Suiza, Holanda y Canadá) y países que presentan ciertas similitudes con España en cuanto a su cultura político-administrativa (Alemania y Francia). España presenta, en comparación con estos países, algunos retos singulares; sin embargo, comparte con ellos otros desafíos, como impulsar la utilización de las evaluaciones en la elaboración de las políticas públicas.

Palabras clave: evaluación, políticas públicas, institucionalización, comparación internacional.

Abstract

This article analyses the situation of public policy evaluation in eight countries, including Spain. Specifically, it explores the institutionalisation of evaluation in the political-administrative system and, in particular, its regulatory and organizational framework, the evaluation capacity of public administrations and the use public decision-makers make of evaluation. The countries were chosen on the basis of various criteria: a long tradition of evaluation (United Kingdom and the United States), countries with a high degree of institutionalisation of evaluation in their political systems (Switzerland, the Netherlands and Canada) and countries with certain similarities to Spain in terms of their political-administrative culture (Germany and France). Compared to these countries, Spain presents some unique challenges; however, it shares with them some common issues such as the need to enhance the use of evaluations in policy making.

Keywords: evaluation, public policies, institutionalisation, international comparison.

JEL classification: H10, H50, H70.

SIMILITUDES Y DIFERENCIAS EN LA EVALUACIÓN DE POLÍTICAS PÚBLICAS EN OCHO PAÍSES: ESPAÑA EN PERSPECTIVA COMPARADA

José María CASADO

Eloísa DEL PINO

AIReF

I. INTRODUCCIÓN: ¿HAN CONVERGIDO LOS SISTEMAS DE EVALUACIÓN DE POLÍTICAS PÚBLICAS?

ALGUNOS análisis han encontrado cierta tendencia convergente en las prácticas y los enfoques de evaluación de políticas públicas (EvPP) en el mundo, pero también han mostrado la persistencia de diferencias en los arreglos y formas de institucionalización de la evaluación entre los países (Stockmann, Meyer y Szentmarjay, 2022; OCDE, 2021; AIReF, 2021; De Francesco y Pattyn, 2021; Stockmann, Meyer y Taube, 2020; Jacob, Speer y Furubo, 2015).

La convergencia entre enfoques de evaluación en los países funciona como un proceso de sedimentación. Conviven los enfoques anteriores con otros más actuales. Los que configuran la denominada tercera generación siguen considerándose los principales enfoques de EvPP en muchos países y en ellos coexisten con la cuarta y quinta generación. En la tercera, el evaluador actúa como un juez, emitiendo juicios sobre la eficacia y la eficiencia. En la cuarta, caracterizada por un enfoque constructivista y pluralista, la voz de los actores es fundamental.

Y en la quinta, el énfasis de la evaluación se pone en el establecimiento de causalidad entre los programas públicos y sus resultados e impactos (Guba y Lincoln, 1990, pp. 38-39; Brousselle y Buregeya, 2018; De Francesco y Pattyn, 2021).

Esta cierta homogenización en los países desarrollados se debe principalmente a dos factores. Uno de ellos es la extensión de la nueva gestión pública (NGP) como paradigma dominante en las administraciones de los países desarrollados. La NGP habría contribuido a que algunos países, que hasta ese momento no evaluaban o lo hacían de forma excepcional, vieran la idoneidad de incorporar esta práctica. La descentralización y desconcentración típicas de la NGP incrementaron la necesidad de coordinación y obtención de información y esta necesidad creció en un contexto de control del gasto público. Ello estimuló evaluaciones de tipo *benchmarking* y la medición cuantitativa del rendimiento (Vedung, 2010, p. 272).

El segundo factor es el esfuerzo de las organizaciones internacionales (UE, OCDE o Banco Mundial) por promover estándares de evaluación (Pattyn *et al.*, 2018). En Europa, países con tradición evaluadora como los

Escandinavos, Reino Unido o Alemania reforzaron su actividad y otros iniciaron sus prácticas evaluadoras tras la adhesión a la Unión Europea (UE). Fue el caso de los países del sur de Europa en la década de 1980 y de los países del Este posteriormente, cuando una gran parte del presupuesto de la UE se redirigió a estos nuevos Estados miembros a través de los fondos estructurales y la política de cohesión. Desde 1988, la Comisión Europea comenzó a exigir evaluaciones. A partir de 1990 reforzó sus requerimientos de evaluación *ex ante* y evaluaciones intermedias a las autoridades nacionales. En los primeros años del siglo XXI introdujo la evaluación de impacto regulatorio que parecía encajar bien con los nuevos instrumentos de política pública (por ejemplo, la corregulación, el método abierto de coordinación, los partenariados y la autorregulación), más complejos y para los que la evaluación convencional no es de fácil aplicación (Stern, 2009; Blomeyer y Eussner, 2020). Finalmente, además de la OCDE, también la UE, en particular a través de sus *joint research centres*, ha tenido cierta influencia en la aplicación de las ciencias del comportamiento y en el hecho de que al menos diez países de la UE, entre los que no está España, cuentan con unidad de *nudges* (John, 2019; De Francesco y Pattyn, 2021).

Sin embargo, a pesar de esta convergencia sedimentaria en la que se acumulan enfoques de evaluación, los arreglos y formas de institucionalización de evaluación no parecen haber convergido de la misma manera (Stockmann, Meyer y Taube, 2020). Aunque la NGP y las instituciones internacionales han podido tener cierto impacto, estas

presiones no han sido iguales ni han tenido el mismo efecto en todos los países (Stern, 2009). Por ejemplo, dado que la evaluación se relaciona con los fondos estructurales, los países receptores de más fondos han recibido más presión y es posible que en ellos las políticas más vinculadas a esta financiación (por ejemplo, el desarrollo regional) hayan concentrado, al menos al principio, la mayoría de las evaluaciones a diferencia de otros países donde la evaluación presenta más diversidad sectorial. Stern (2009) también sugiere que, en los países que han recibido más financiación europea, es más probable que se desarrolle una política centralizada de evaluación sobre todo si esta se vincula al control del gasto a través de los ministerios de hacienda (especialmente en los países del Este) en contraste con otros países como el Reino Unido o Suecia y Holanda, con una cultura descentralizada de evaluación. Finalmente, aunque a veces los estímulos que reciben los países de las organizaciones internacionales son similares, estos han pasado por el filtro de las instituciones y la cultura administrativa nacional produciendo rutas distintas en el proceso de institucionalización de la EvPP.

En este artículo tratamos de identificar similitudes y diferencias en la institucionalización de la EvPP en los países de nuestro entorno dentro del nivel central de Gobierno y ubicar el caso de España en este contexto internacional. En particular comparamos: 1) la existencia de regulación general, o al menos sectorial, en la que se establezcan las características de la política de EvPP y en identificar un marco organizativo en el que se

desarrolla; 2) la capacidad de las administraciones para hacer EvPP, el tipo de evaluación que se practica, su naturaleza interna o externa, si existen guías o mecanismos para velar por la calidad de las evaluaciones; y 3) el uso que se hace de la EvPP por parte del Parlamento o del Ejecutivo. Estas cuestiones indican hasta qué punto el sistema político-administrativo de un país ha hecho un esfuerzo por institucionalizar la EvPP (Stockmann, Meyer y Taube, 2020). Además, estos asuntos atañen a un buen número de las preocupaciones que han mostrado los expertos españoles en EvPP (Bustelo, 2020; Onrubia y Fernández-Fuentes, 2019; Ruiz, 2012; Garde-Roca, 2017; Catalá *et al.*, 2015; De la Rica, 2020; Herrero, 2021; Fedea, 2021; Jimeno, 2021).

Para responder a estas preguntas, se realiza una revisión de la literatura existente sobre EvPP y se analizan numerosos documentos y webs sobre evaluación de cada país estudiado. Se han escogido ocho países atendiendo a diversos criterios: países con larga tradición de evaluación (Reino Unido y Estados Unidos), países que presentan ciertas similitudes con España en cuanto a su cultura administrativa (Alemania y Francia) y países que han avanzado más en el proceso de institucionalización de la EvPP en sus sistemas político-administrativos (Suiza, Holanda y Canadá) (Stockmann, Meyer y Taube, 2020; Stockmann, Meyer y Szentmarjay, 2022). También se aprovecha la experiencia laboral de los autores en instituciones responsables de evaluación en España.

El resto del artículo se organiza en cuatro secciones. En la siguiente se analiza el marco

normativo y organizativo de la EvPP en los países seleccionados. En la tercera sección se revisa lo que hemos llamado la capacidad de evaluación para referirnos a la disponibilidad de recursos de evaluación en las administraciones de cada país. En el cuarto apartado se hace referencia a la difusión y la utilización de la EvPP. En la sección final se discuten los hallazgos.

II. LA POLÍTICA DE EVALUACIÓN: MARCO ORGANIZATIVO Y NORMATIVO DE LA EVALUACIÓN

Desde 2000, la Constitución federal suiza establece que el Parlamento velará por que las medidas del Gobierno federal sean evaluadas para estimar su eficacia. Aunque ya desde la década de 1990 se identifican importantes iniciativas, la referencia constitucional a la evaluación contribuyó a un nuevo impulso en el país (Widmer, 2000; Horber-Papazian y Baud-Lavigne, 2019). Fruto de la reforma constitucional, en 2002, un grupo interministerial elaboró un informe con recomendaciones de aplicación en todos los órganos de la Administración y el Parlamento. Wirths y Horber-Papazian (2016) encontraron que el 80 por 100 de las disposiciones sobre EvPP de la legislación federal y cantonal se incluyeron después de la reforma constitucional. La mayoría de ellas recogen la necesidad de realizar evaluaciones de tipo retrospectivo y sectorial, lo que ha ocasionado que algunos expertos sugieran una normativa más transversal que permitirá economías de escala (Widmer, 2020).

Desde 1990, el Parlamento federal cuenta con su propia unidad de evaluación interna deno-

minada Control Parlamentario de la Administración (CPA). La CPA realiza evaluaciones de forma continuada sobre la legalidad y la eficacia de las actividades del Gobierno federal en nombre de las comisiones de control de las dos cámaras parlamentarias. El Ejecutivo está obligado a responder a las recomendaciones. La CPA coordina sus planes de evaluación con el Tribunal de Cuentas que también evalúa la relación coste-beneficio de las medidas de la Administración y sus efectos. Sin embargo, en términos numéricos, las evaluaciones que realizan estos órganos son mucho menos numerosas que las 80 anuales que elaboran o encargan los ministerios suizos. Algunas agencias federales evalúan con frecuencia (la Oficina de Salud Pública, la Agencia para la Cooperación al Desarrollo, la Secretaría de Estado de Asuntos Económicos, la de Educación y la de Investigación e Innovación), pero también hay otras que nunca han evaluado (Horber-Papazian y Baud-Lavigne, 2019; Widmer, 2020).

Holanda es, junto a Suiza, uno de los países europeos de entre los once que analizan Stockmann, Meyer y Taube (2020) con un alto grado de institucionalización de la EvPP en su sistema político-administrativo. La evaluación comenzó a impulsarse en la década de 1970 y ha florecido a partir de los 2000, cuando el país ha formalizado una política de evaluación (Klein Haarhuis, 2020, p. 90). La evaluación está regulada en la Ley de Cuentas del Estado de 2001, que ha experimentado modificaciones a lo largo del tiempo, como la realizada en 2018 para exigir que todas las propuestas de política pública incluyan claramente sus objetivos y los instrumentos que usarán para conseguirlos, de

forma que se logre fundamentar mejor las políticas y garantice su evaluación. También existe un reglamento sobre estudios de evaluación periódica de 2012 (Klein Haarhuis, 2020: p. 90). Además, las disposiciones sobre evaluación en normas sectoriales son frecuentes. En algunos sectores, como infraestructuras, la evaluación *ex ante* es obligatoria.

Los ministros son responsables de evaluar periódicamente sus políticas, lo que no significa que todos lo hagan de igual modo. El Tribunal de Cuentas también analiza la eficacia y la eficiencia de las políticas y el Parlamento puede determinar si las propuestas que llegan del Gobierno están bien fundamentadas y si las evaluaciones son suficientes (Klein Haarhuis, 2020, p. 90). En Holanda, el CPB, Oficina de Análisis de Política Económica, una agencia administrativa independiente, también realiza evaluaciones, incluidas algunas sobre las medidas propuestas por los partidos políticos, a petición de estos (1).

Las dos principales formas de evaluación presupuestaria son los *spending reviews* (SR) y las *policy reviews* (de Jong, 2016). Los primeros han ido evolucionando hacia evaluaciones destinadas a comprobar la eficacia y eficiencia e identificar alternativas de política pública y acostumbran a proponer posibles ahorros en el gasto. El Ministerio de Hacienda presenta anualmente al Consejo de Ministros la propuesta de *spending*. De acuerdo con de Jong (2016), en tres décadas se habrían llevado a cabo alrededor de 270 SR. Además, desde 2006, los ministros son responsables de realizar *policy reviews* periódicamente (cuatro a siete años) que deben revisar la teoría del cam-

bio o fundamentos subyacentes a las intervenciones públicas. Recientemente se ha puesto en marcha una nueva forma de evaluación, inspirada en los *public value reviews* del Reino Unido, los *publieke waardescan*, que revisan hasta qué punto se han tomado las medidas oportunas para maximizar el valor añadido esperado de la política analizada (Ministerio de Hacienda de Holanda, 2021). Además, también llevan a cabo evaluaciones de coste beneficio y de otros tipos.

Finalmente, desde 2014 cada ministerio debe tener su propio equipo de análisis del comportamiento y el Ministerio de Economía se encarga de homogeneizar sus iniciativas y evaluaciones a través de la *Behavioural Insights Network Netherlands* (Banco Mundial, 2019).

Canadá cuenta con una larga tradición de evaluación que se remonta a 1977, cuando elaboró su normativa *Policy on Evaluation*. Este documento se actualiza periódicamente incluyendo nuevos enfoques e iniciativas (Gauthier, Lahey y Jacob, 2022). Desde 2016, con la denominación de *Policy on Results* y, junto con una directiva con el mismo nombre, es la norma clave que inspira la EvPP en el nivel federal. Entre los cuatro objetivos que espera conseguir enumera que los «ministerios midan y evalúen su rendimiento y usen esta información para gestionar y mejorar los programas, políticas y servicios» (2). En 2001, reguló la obligatoriedad de la evaluación en los ministerios, lo que no significa que ya sea una práctica generalizada.

Su marco *Policy on Results* ha llevado a la creación de unidades de EvPP en la mayoría de los mi-

nisterios y se han puesto en marcha comités de seguimiento de la evaluación. Cada ministerio es responsable de formalizar un plan de evaluación y los programas que superen los cinco millones de dólares canadienses deben evaluarse cada cinco años. Además, debe disponer de un jefe de evaluación y un comité de medición y evaluación del desempeño.

Sin embargo, es el Consejo del Tesoro (TBS) quien tiene la misión de extender la cultura y coordinar la política de evaluación. De hecho, los ministerios deben obtener la aprobación del TSB para desarrollar o modificar su marco (*Departmental Results Framework, DFR*) y estructuras de evaluación. El *DFR* se sintetiza en un cuadro en el que se incluyen: 1) responsabilidades básicas de un ministerio; 2) los objetivos e indicadores; y 3) un inventario de programas que contribuyen a garantizar las competencias básicas y resultados del departamento (3). El Parlamento es el destinatario de todas las evaluaciones realizadas por los ministerios. Además, en 2015, Canadá creó la Unidad de Impacto e Innovación (IIU), dependiente del primer ministro y que trabaja con los ministerios en el desarrollo de las metodologías de medición del impacto y en evaluar los efectos de los experimentos.

Por lo que se refiere al Parlamento, la Oficina de Presupuestos (OPC) proporciona a los diputados análisis independientes sobre el presupuesto o sobre otros temas relacionados con la economía nacional. También puede estimar el coste de cualquier propuesta legal si lo solicita una comisión parlamentaria. Aun así, la OPC no tiene un mandato específico en materia de evaluación (OCDE, 2021).

Por otro lado, las distintas comisiones parlamentarias pueden pedir informes a la Oficina del Contralor General de Canadá (OCG) con el fin de supervisar las actividades del Gobierno y evaluar el gasto público. Realiza una importante labor de auditoría de distintos tipos, incluyendo en algunas de ellas evaluaciones del impacto de los programas públicos de diferentes sectores.

Alemania se interesa por la EvPP desde la década de 1960, cuando implementó una profunda reforma de su Administración Pública. Más tarde, en los años 1990, debido a la influencia de la NGP y de la UE, y en la década de 2000, como resultado de las grandes reformas del mercado laboral, la EvPP experimentó su auge. En Alemania no hay una legislación ni una política nacional sobre evaluación, aunque, desde la década de 1970, su Código Presupuestario Federal dispone que las auditorías deben valorar el logro de los objetivos, los impactos y la rentabilidad. El Reglamento Interno Conjunto de los Ministerios (que regula la organización de los ministerios y la cooperación entre ellos y con los órganos constitucionales federales) también exige la realización de la evaluación de impacto y el análisis coste-eficacia. Además, la normativa sectorial que recoge obligaciones concretas de evaluación ha aumentado significativamente en las últimas legislaturas, lo que no quiere decir que estas obligaciones se cumplan siempre (Stockmann y Meyer, 2020).

De entre los analizados en este artículo, Alemania es quizá el país con un sistema de evaluación a nivel federal menos coordinado, en el que los ministerios gozan de gran autonomía (OCDE, 2021). Algunos minis-

terios cuentan con unidades especializadas de evaluación y la practican de forma habitual (por ejemplo, el ministerio encargado de la cooperación al desarrollo). Otras veces, los ministerios se apoyan en entidades independientes a las que financian en gran medida. Es el caso de la agencia de exportación del Ministerio de Economía, que cuenta con una unidad que realiza evaluaciones o supervisa las que hace el ministerio o las que este encarga a terceros. Existe también un Instituto de Investigación sobre el Empleo dependiente del Servicio Público de Empleo que se encarga de las evaluaciones y cuenta con numeroso personal y un elevado presupuesto (Consejo de Estado francés, 2020).

En el Parlamento, el Tribunal de Cuentas realiza una extensa labor de auditoría del Gobierno federal y otras actividades que reflejan crecientemente su interés por la EvPP como puede comprobarse en el informe 2020 (4). Su gran estructura y plantilla le permiten realizar actividades de asesoramiento y formación. Finalmente, desde 2015, Alemania creó un equipo de ciencias sociales y del comportamiento con el apoyo directo de la Cancillería. La unidad trabaja con los ministerios para diseñar y aplicar intervenciones y algunos (defensa o empleo) ya han evaluado programas diseñados con este enfoque.

En el Reino Unido, la profesionalización de la gestión pública a partir de 1970 y el deseo de control de gasto público fueron las dos principales fuerzas impulsoras de la evaluación (Pattyn *et al.*, 2018). Desde entonces, se han sucedido distintos programas que han ido poniendo el énfasis en diferentes tipos de evaluaciones y metodologías,

pero no existe una regulación general sobre evaluación. Algunos nuevos programas incluyen disposiciones más o menos vinculantes para su evaluación. Sin embargo, el Ministerio de Hacienda y el *Cabinet Office* (lo que aquí sería el Ministerio de Presidencia) sí desempeñan funciones de dirección informal de las evaluaciones (OCDE, 2021). En todo caso, cada ministerio es responsable de planificar y ejecutar sus propias evaluaciones.

Además, existen otros organismos públicos no vinculados a los ministerios que también desempeñan tareas relacionadas con la evaluación. La Oficina Nacional de Auditoría (NAO) es un organismo parlamentario independiente que desarrolla una importante actividad de supervisión del gasto en todos los sectores de política pública. También hay otros organismos que tienen un papel en materia de evaluación, como el Instituto de Estudios Fiscales (IFS), en concreto en el análisis presupuestario de los proyectos de ley de financiación o las propuestas de los principales partidos políticos en períodos preelectorales.

En 2010, se creó *Behavioural Insights Team (BIT)* con la finalidad de tener más en cuenta la investigación en ciencias del comportamiento en el diseño y EvPP y para ello realizó experimentos e informes conjuntos con los principales ministerios. Su informe de 2012, *Test, Learn, Adapt* (5), tuvo una amplia difusión y dio a conocer los métodos y técnicas experimentales (RCT) en la Administración. Desde 2014, el *BIT* es independiente, aunque continúa siendo en parte financiado por el Gobierno y asesorando a las administraciones (Baiz *et al.*, 2019) (6). Durante el Gobierno de Tony Blair

se crearon los centros *What Works* que tuvieron una gran influencia en el uso de la evidencia (Baiz *et al.*, 2019). Desde 2015, el equipo de *What Works* coordina el denominado *Cross-Government Trial Advice Panel (CG-TAP)*, compuesto por académicos y expertos de la Administración con el fin de estimular la evaluación en los ministerios. Aunque no realiza evaluaciones, ha apoyado 52 de ellas (7).

En Estados Unidos, la evaluación se extendió con la expansión de los programas sociales de la *Great Society* y, en concreto, con la exigencia del Gobierno federal para que los estados evaluaran las políticas educativas y de protección social que aquel financiaba. El país cuenta con un marco formal de evaluación desde hace tiempo. Sin embargo, desde 2018, debe destacarse el impulso que ha supuesto la Ley de Fundamentos para la Elaboración de Políticas Basadas en la Evidencia (Ley de Evidencia) (8) que insta al Gobierno federal a tomar decisiones utilizando la mejor evidencia disponible a través de un análisis sistemático orientado por el principio de integridad científica. Además, en esta fecha, el Gobierno puso en marcha un plan de reforma para el fortalecimiento de la evaluación en este nivel federal (9). En 2021, a través de un Memorando sobre el Restablecimiento de la Confianza en el Estado a través de la integridad científica y la elaboración de políticas basadas en la evidencia (10), el presidente ha comunicado a los altos cargos de los ministerios la urgencia de aplicar el enfoque de políticas basadas en la evidencia, para lo cual se organizó un grupo cuyo trabajo ha dado lugar a variadas medidas que afectan también a la EvPP y que se mencionan más adelante.

Desde 2018, la Oficina de Gestión y Presupuesto (OMB) de la Casa Blanca cuenta con un pequeño equipo, el *Evidence Team*, encargado del impulso de la evaluación y su coordinación en la Administración federal, aunque los ministerios pueden decidir las evaluaciones que realizan y que incluyen en sus planes. Algunos ministerios cuentan con potentes unidades de análisis y evaluación, por ejemplo, el Ministerio de Trabajo y su oficina CEO (11).

El Congreso de los Estados Unidos dispone de tres organismos que realizan actividades de evaluación, aunque no necesariamente de impacto. La General Accounting Office (GAO) cuenta con un importante presupuesto y realiza fundamentalmente auditorías y análisis de la eficacia y eficiencia. Elabora entre 600 y 700 análisis al año y el seguimiento de sus recomendaciones se puede consultar en línea (OCDE, 2021) (12). Además de ella, el Congreso cuenta un Servicio de Investigación (CRS) que realiza análisis legislativos y una Oficina Presupuestaria (CBO). Esta última, elabora proyecciones económicas y presupuestarias y estimaciones de coste de los proyectos de ley y otras normas.

En Francia se pueden identificar algunas experiencias de evaluación en diversos ministerios desde la década de 1960, pero la EvPP ha sufrido altibajos a lo largo del tiempo. Desde 2008, la constitución francesa mandata al Tribunal de Cuentas que asista al Parlamento y al Gobierno en la EvPP (Consejo de Estado de Francia, 2020). En 2009, se creó el Comité de Evaluación y Control (CEC), que coordina una media de seis informes de evaluación cada año. A pesar

de todo, el papel del Parlamento en la EvPP es limitado, aunque ha comisionado algunos estudios emblemáticos a cuatro centros universitarios y en 2019 contaba con un presupuesto para encarar otros.

En Francia no existe un organismo de coordinación ni una política de EvPP a nivel nacional. Cada ministerio programa su propio plan de trabajo. Aun así, cabe mencionar el programa de evaluación que el Gobierno propuso en 2012 en el marco de la iniciativa de Modernización de la Acción Pública (MAP). Esta iniciativa dio lugar a más de 80 informes de evaluación (Lacouette-Fougère y Simon, 2020). Además, France Stratégie, organismo de prospectiva dependiente del primer ministro, creado en 2013 y sucesor del Centre d'Analyse Stratégique, ha asumido tareas de evaluación que representan el 20 por 100 del total de sus actividades (Consejo de Estado francés, 2020). También han surgido otros actores, en particular centros de evaluación económica, que son responsables de alrededor del 15 por 100 de las evaluaciones. Según la Sociedad Francesa de Evaluación, entre 2007 y 2017, se produjeron unas 236 evaluaciones por año, un 45 por 100 de las mismas por parte del Estado central. Del conjunto de las evaluaciones, un 12 por 100 son *ex ante*, un 43 por 100 durante y un 46 por 100 *ex post*; en un 50 por 100 se aplicó una metodología mixta o cuantitativa (13).

Por último, en 2013, la Secretaría General para la Modernización del Estado (SGMAP) puso en marcha algunos proyectos con enfoque experimental en diversos sectores de política pública. Más tarde, estas responsabilidades se situaron en una

dirección general que actualmente se ubica en el Ministerio de Transformación y Función Pública (14).

En España, la EvPP que se llevó a cabo entre 1980 y 1990, se concentró en unos pocos sectores (educación, empleo, servicios sociales, sanidad, fondos europeos o cooperación al desarrollo) y no tenía carácter sistemático (Fenstein y Zapico-Goñi, 2010; Bustelo, 2020). Este despegue tardío empieza a corregirse a partir del ingreso en la UE. Desde entonces pueden destacarse otros dos hitos. El primero fue la creación de la Agencia Estatal de Evaluación de Políticas Públicas y Calidad de los Servicios (AEVAL) en 2006. Sin embargo, a partir de 2011, la AEVAL fue perdiendo su impulso inicial y en 2017, en un contexto de racionalización del sector público, se disolvió y se creó en su sustitución el Instituto para la Evaluación de Políticas Públicas (IEPP) con rango de subdirección general adscrita a la Secretaría de Estado de Función Pública.

El segundo hito fue la asunción de tareas y competencias de evaluación por parte de la Autoridad Independiente de Responsabilidad Fiscal (AIReF) que ha proporcionado un fuerte impulso a la EvPP en el país. La AIReF, donde en 2021 se ha creado una División de Evaluación del Gasto Público, se ha convertido en el principal agente evaluador en los últimos años en el nivel central de Gobierno desde que en 2017 realizara su primera evaluación. El Gobierno central encarga periódicamente a la AIReF los llamados *spendings reviews* (el primero abarcó el período 2018-2021 en tres fases y un segundo constará de cinco fases que se desarrollarán en el período 2022-2026) y otras eva-

luaciones concretas (alrededor de doce comunidades autónomas le han encargado también varias evaluaciones). La AIReF se crea en 2013 como respuesta a la crisis de 2008 para el refuerzo de la estabilidad presupuestaria y gracias al impulso europeo, y está regulada en su Estatuto Orgánico de 2014. En este sentido, la puesta en marcha de la evaluación en la AIReF genera importantes sinergias con su función tradicional: su misión de garante de las finanzas públicas se enriquece ahora con los análisis de la calidad del gasto público. Su independencia se garantiza por, entre otros mecanismos, su forma de financiación mediante una tasa y el modo de selección de su presidenta, que requiere la mayoría absoluta del Parlamento y se realiza fuera de las legislaturas.

Además de la AIReF, en la Administración central existen otros organismos que realizan evaluaciones. Es el caso del Instituto de Estudios Fiscales o el mencionado IEPP. Algunos ministerios realizan internamente evaluaciones o las subcontratan con terceros. Finalmente, el Tribunal de Cuentas (vinculado al Parlamento) y la Intervención General de la Administración del Estado (dependiente del Ejecutivo) realizan tareas de fiscalización y auditoría pública y otras actividades cercanas a la EvPP.

Aunque en España no existe una ley de EvPP, Onrubia-Fernández y Sánchez-Fuentes (2019, p. 236) detectan «una presencia... extensiva y hasta bastante exagerada del término evaluación en nuestra legislación». Recientemente, por ejemplo, la evaluación de los resultados de las políticas se menciona en la Ley de Régimen Jurídico del Sector Público de 2015. Además,

la propia Constitución se refiere a la asignación equitativa del gasto público y a que su programación y ejecución responderán a los criterios de eficacia y eficiencia. Igualmente, la Ley Orgánica de Estabilidad Presupuestaria y Sostenibilidad Financiera establece el principio de eficiencia en la asignación y utilización de los recursos públicos. Sin embargo, estas menciones no han garantizado la práctica de la EvPP. Tampoco es claro que todas estas referencias aludan a actividades que puedan considerarse como EvPP y no como ejercicios de fiscalización o auditoría. En cualquier caso, en 2021, el Gobierno de España abrió una consulta pública para la redacción de un anteproyecto de ley y el Plan Normativo nacional para 2022 incluye la aprobación de una ley de evaluación en el último trimestre de 2022 (15).

Además, el Plan de Recuperación, Transformación y Resiliencia (PRTR) presentado por el Gobierno en 2021 para acceder a los nuevos instrumentos comunitarios de financiación *Next Generation EU*, menciona la EvPP en varios de sus componentes (muy concretamente en el 11 sobre Modernización de las Administraciones Públicas; y en el 29 sobre Mejora de la Eficacia del Gasto Público). Por un lado, consolida el papel que tendrá la AIReF como organismo evaluador reforzando su independencia, profesionalidad, sistematicidad y la relevancia de sus evaluaciones (16).

Además, en los últimos meses de 2021 y en 2022 algunos anteproyectos de ley están incorporando en su articulado compromisos relevantes en materia de evaluación (creación de estructuras de evaluación, evaluación

periódica, entre otros), como la Oficina del Análisis del Empleo recogida en el Anteproyecto de la Ley de Empleo. También cabe destacar la puesta en marcha de proyectos piloto innovadores, como los financiados por el Ministerio de Inclusión, Seguridad Social y Migraciones y ejecutados por las administraciones subcentrales alrededor del ingreso mínimo vital y su evaluación, que debe estar lista en 2024 (17). En el desarrollo y evaluación de estos itinerarios de inclusión, el Ministerio de Inclusión colabora con una organización nacional, el Cemfi, vinculada al Banco de España, y otra internacional, J-Pal, con experiencia en ensayos de control aleatorio (18).

III. LA CAPACIDAD EVALUADORA: RECURSOS Y EXPERIENCIA

En Suiza, la Sociedad de Evaluación ha tenido una gran influencia en el desarrollo de la evaluación, en particular desde 2001, cuando realizó una importante labor en la formalización de los estándares de evaluación que se basaron en los de Estados Unidos. Aunque el número de personas dedicadas a la evaluación en la Administración federal ha crecido en los últimos años (Horber-Papazian y Baud-Lavigne, 2019), la mayor parte de las evaluaciones se contratan externamente y el papel de la Administración y de la académica en su ejecución ha disminuido. Cada vez más, las administraciones se limitan a diseñar los proyectos de forma estandarizada. Las consultoras ya establecidas invierten crecientes recursos en la elaboración de los proyectos con los que se presentan a las muy competidas licitaciones para asegurar que

pueden conseguirlos. Ello hace que, cuando lo ganan, destinen menos recursos a la ejecución del proyecto dado el trabajo realizado *ex ante*. Además, en los últimos años se ha detectado un incremento de la burocratización del proceso que ha contribuido a que las evaluaciones tengan un cariz más legalista y se parezcan más a las auditorías (Widmer, 2020). El país discute si sería necesario diseñar un mecanismo de acreditación de los evaluadores. Finalmente, los recursos financieros asignados a la evaluación en el nivel federal se han estancado durante muchos años (Horber-Papazian y Baud-Lavigne, 2019).

En Holanda, en los ministerios no hay unidades de evaluación como tales. La mayoría de las evaluaciones se contratan con consultoras, con la academia o con institutos de investigación semiindependientes. El porcentaje que se hace internamente en cada ministerio varía y, en parte, depende de la experiencia (Klein Haarhus, 2020, p. 90). En 2018, el Gobierno puso en marcha la operación *Insight into Quality* dirigida por el Ministerio de Hacienda para mejorar las políticas, desde su diseño a su evaluación de impacto (Ministerio de Hacienda holandés, 2021). Esta iniciativa respondía a las críticas del Tribunal de Cuentas y de un Grupo de Estudio sobre Fiscalidad que consideraban que la información que se obtenía de las políticas no era de calidad suficiente. Durante los años que ha durado la iniciativa, los ministerios debían: identificar los asuntos de política pública que era necesario investigar, mejorar la fundamentación de las políticas que querían proponer y planificar las evaluaciones. Ello debía configurar una Agenda de Evaluación Estratégica (AEE).

En 2018 se pusieron en marcha 19 iniciativas en los ministerios consistentes en experimentos, el desarrollo de sistemas de indicadores en políticas como la de medioambiente o formas innovadoras de analizar los datos. En 2021, se publicó un informe con las conclusiones de esta iniciativa (19).

En el contexto de *Insight into Quality* se han tomado medidas que en parte respondían a la discusión existente entre los expertos sobre la necesidad de establecer estándares de evaluación. La deficiente calidad de los *policy reviews* ya había llevado a desarrollar un instrumento denominado *beleidsdoorlichtingen*, consistente en listado de cuestiones que debe responder de forma estandarizada cada ministerio. Las respuestas son examinadas por un experto externo y enviadas al Parlamento por el Consejo de Ministros. El instrumento se ha revisado en varias ocasiones. Se ha desarrollado un programa de formación y elaborado una interesante caja de herramientas (*toolbox beleidsevaluatie*) que puede consultarse en línea (20).

En Canadá, la mayoría de las evaluaciones de los programas son realizadas internamente por unos 450 evaluadores del Gobierno federal (Baiz et al., 2019). Al crecer la demanda interna de evaluación, algunos evaluadores que trabajaban en firmas privadas han pasado a formar parte de la plantilla de la Administración. Esto ha hecho que se contraten menos proyectos de evaluación completos con el sector privado y se licite prioritariamente el trabajo de campo (Gauthier, Lahey y Jacob, 2022, p. 150). El Consejo del Tesoro (TBS) evalúa y certifica las competencias de los evaluadores, que detalla en

su página web (21) y las capacidades de las unidades de evaluación de ministerios y la calidad de las evaluaciones. La División de Resultados, formada por unos 50 analistas, apoya el desarrollo de unidades de evaluación dentro de los ministerios (OCDE, 2021). Ofrece también formación en las denominadas *drop-in sessions* para los evaluadores y las difunde a través de las redes sociales (OCDE, 2021). Canadá dispone de numerosas guías de evaluación. Entre las últimas, destaca una sobre la evaluación con perspectiva de género (22).

Por último, Canadá también ha desarrollado una hoja de ruta para la explotación de datos de la Administración. El trabajo conjunto del TBS, de Statistics Canada y de los distintos ministerios y la comunidad evaluadora se considera un factor de éxito que ha mejorado el uso de los datos.

En Alemania, las evaluaciones son realizadas principal y regularmente por institutos de investigación no universitarios más o menos vinculados a las estructuras gubernamentales y al mundo académico y a los que la Administración realiza sus encargos. Estos institutos están muy presentes en el debate económico y son también productores de datos estadísticos. Algunos son generalistas (Centro de Investigación en Ciencias Sociales de Berlín [WZB]; Instituto Alemán de Investigación Económica [DIW]); otros están especializados en un tema concreto (Instituto de Economía Laboral [IZA]; Instituto de Investigación sobre el Empleo [IAB]). Algunos son empresas privadas, la mayoría sin ánimo de lucro y en parte con financiación pública. Otros están vinculados por ley a las autoridades administrativas, como el IAB, crea-

do por el Ministerio de Trabajo. Cuentan con una financiación y un personal relativamente importantes, al menos desde la década de 2010. La disponibilidad de datos en algunos sectores, como el del mercado laboral, favorece la evaluación (Baiz et al., 2019).

En Alemania, algunos análisis han mostrado que no existe una falta de voluntad política o administrativa en relación con la EvPP. El problema reside más bien en la capacidad ministerial para realizar evaluaciones. En 2013, un informe del Tribunal de Cuentas demostró que las evaluaciones presentan problemas en cuanto a su alcance, metodología y la *expertise* de los evaluadores (Stockmann y Meyer, 2020). En algunos ministerios se ofrece formación sobre metodología de EvPP, de manera sistemática (Baiz et al., 2019). Aun así, el Tribunal de Cuentas y el Consejo Nacional de Control de la Regulación han amonestado a los ministerios en alguna ocasión por la calidad de sus evaluaciones.

En la actualidad, Alemania no dispone de directrices generales sobre cómo llevar a cabo evaluaciones de impacto o supervisar la eficacia de los programas, pero se están elaborando (23). Algunos ministerios han publicado orientaciones internas. El Consejo Nacional de Control de la Regulación (NKR), como órgano consultivo independiente, ha desarrollado un modelo de buena evaluación que puede utilizarse como posible base para las normas de evaluación de los ministerios federales. La Sociedad Alemana de Evaluación ha publicado unos estándares de evaluación que se espera que puedan contribuir a mejorar la calidad de las evaluaciones.

El Gobierno del Reino Unido realiza un esfuerzo intenso para la mejora de la capacidad evaluadora interna, aunque con frecuencia las evaluaciones se subcontratan con consultoras o centros de investigación buscando independencia. Existen tres iniciativas destacables. La primera fue la creación en 2004 de la *Policy Profession*. Para este grupo de empleados públicos se exigen competencias en evaluación que se describen en 23 criterios, entre los que se encontraban algunos como: comprender la importancia de evaluar las políticas, utilizar evaluaciones anteriores para analizar la eficacia de las políticas; capacidad para criticar una evaluación desde un punto de vista metodológico e identificar estrategias de investigación; experiencia práctica en una serie de métodos de evaluación o capacidad para encargar evaluaciones y apoyar su uso en el proceso de toma de decisiones (Baiz et al., 2019). Estos criterios se han utilizado para la selección de un volumen de personal que en 2017 ascendía a casi 18.300 personas. También existe formación continua para los departamentos ministeriales.

Desde 2015, el equipo de *What Works* anima a los ministerios a publicar las preguntas sobre las que les gustaría conseguir evidencia (Áreas de Interés para Investigar, ARI) (Boaz et al., 2021). En 2022, se ha publicado un guía para elaborar los (24). Además, el Ministerio de Hacienda actualizó en 2020 los *Libros Verde y Magenta* (25) sobre evaluación que se utilizan como guías de EvPP y reflexionan sobre las capacidades necesarias para evaluar. El *Libro Magenta* cuenta, con una adenda metodológica de 2020, en el que se presentan 30 técnicas

cuantitativas y cualitativas aplicables a la evaluación (26). En 2019 se actualizó el *Public Value Framework* (27) que ayuda a que los ministerios revisen las políticas en marcha para detectar posibles asuntos a mejorar.

En Estados Unidos, los ministerios pueden desarrollar sus evaluaciones o contratarlas con terceros. La Ley de la Evidencia introdujo la llamada Agenda de Aprendizaje (AA) que debe formar parte del plan estratégico de la agencia o ministerio e identificar qué temas de política pública necesitan evidencia y cómo se abordará su análisis, incluyendo la metodología que se utilizará. Existen numerosas guías para elaborar la AA. En particular, la Oficina de Ciencias de la Evaluación (OES) ha desarrollado un interesante material de apoyo (28), tanto para la elaboración de las AA como para el diseño de los planes de evaluación y el análisis de las capacidades de los ministerios. La oficina cuenta con 30 efectivos, entre los cuales se encuentran investigadores de numerosas disciplinas que proporcionan apoyo al Gobierno para el diseño y EvPP utilizando un enfoque de ciencias del comportamiento. Entre 2015 y 2022 han llevado a cabo más de 90 evaluaciones de impacto en 20 agencias gubernamentales de distintos sectores.

Tras la aprobación de la Ley de Evidencia de 2018, el Gobierno está realizando un intenso esfuerzo por mejorar la capacidad de los ministerios para obtener evidencias, cuya situación es muy desigual (Donaldson, Donaldson y Renger, 2022). Cada ministerio debe evaluar su capacidad para trabajar con evidencia y proponer formación u otras estrategias para mejorarla. Además, la Ley

de Evidencia requiere que los ministerios diseñen un Plan de Evaluación Anual y designen a un responsable de evaluación, cuya cualificación y funciones se establecen con detalle (véase OMB, 2021, pp. 29 y ss.). El responsable de evaluación diseña y aplica la política de evaluación y supervisa las evaluaciones y es el enlace con la OMB de la Casa Blanca. En 2019, la OMB creó un consejo interdepartamental para apoyar el desarrollo de estas capacidades en todo el Gobierno federal. La Ley de Evidencia prevé que la Oficina de Gestión de Personal (OPM) identifique las competencias necesarias para evaluar programas y defina la cualificación de los profesionales.

Finalmente, cabe destacar que la Ley de Evidencia se diseñó en parte para garantizar la disponibilidad y calidad de los datos administrativos en las evaluaciones. Cada agencia debe nombrar un órgano de gobernanza del dato que será responsabilidad de un jefe de datos nombrado entre el personal cualificado de la agencia. Además, debe haber una responsable de estadísticas. La Ley de Evidencia incorpora la Ley de Datos Abiertos del Estado, que obliga a los organismos a publicar información en línea usando formatos reutilizables (OMB, 2021).

En Francia, los servicios estadísticos de algunos ministerios utilizan los datos que producen para hacer evaluaciones. Las áreas con más evaluaciones son también las que producen más datos: economía, sanidad, servicios sociales y empleo (Lacouette-Fougère y Simon, 2020). Una interesante metaevaluación de 80 informes realizados entre 2012 y 2017 detectó buenas prácticas relativas al

planteamiento de los objetivos en las evaluaciones, pero también algunas carencias metodológicas, de recogida de datos y problemas de replicabilidad que en parte fueron corrigiéndose con el apoyo de la Secretaría General para la Modernización de la Acción Pública y con el propio aprendizaje de los evaluadores. Algunos de ellos tenían experiencia en auditorias o en inspección de servicios, pero no en evaluación (Lacouette-Fougère y Simon, 2020). Según el Consejo de Estado francés (2020), el Ministerio de Acción y Cuentas Públicas promueve la evaluación y proporciona apoyo financiero y metodológico a los ministerios. De acuerdo con la Sociedad francesa de Evaluación, casi el 65 por 100 de las evaluaciones se han encargado a consultoras privadas, un 3 por 100 a las universidades y el resto, un 27 por 100, a distintos órganos de la Administración.

En los últimos años, France Stratégie ha elaborado numerosas guías metodológicas y ha organizado seminarios dedicados a la evaluación. También se ha trabajado en la mejora de disponibilidad de información con la creación del Centro de Acceso Seguro a los Datos en el que están involucrados varios ministerios y centros de investigación (29).

En España, salvo algunas unidades especializadas, la capacidad para hacer evaluaciones de los ministerios es limitada debido a la falta de experiencia. Por su lado, los organismos encargados de evaluar cuentan con un número de efectivos reducido (el IEPP cuenta con once, y en la AIReF el personal dedicado a la evaluación no asciende a una decena). La AIReF busca apoyo

externo (consultoras y/o universidades) para la mayoría de las evaluaciones, pero se involucra intensamente con cada uno de los equipos externos, diseñando y dirigiendo las evaluaciones y coordinando su gobernanza. El IEPP ha desarrollado numerosas guías de evaluación, pero su actividad evaluadora de los últimos años es limitada. Ambas instituciones imparten seminarios de formación. Por su parte, desde hace algunos años existe el Grupo de Referencia en Evaluación de la Administración General del Estado que se reúne mensualmente.

Uno de los asuntos que más se ha discutido en los últimos años en España ha sido la disponibilidad de datos para la realización de las evaluaciones. En este sentido, la AIReF (2020) propuso el diseño de una estrategia nacional (30) que fue compartida también por la comunidad científica e investigadora (Almunia y Rey Biel, 2021) y que debería requerir avances legales para que los datos se produzcan, además de con fines estadísticos, con fines de investigación y evaluación de políticas públicas a la vez que se crean uno o varios repositorios de la información que sean accesibles por las instituciones y la academia. En los últimos años se han producido algunos avances, como la puesta en marcha del Plan de Gobierno Abierto, la creación de la Oficina del Dato para dinamizar la compartición y el uso de los datos en todos los sectores productivos; se ha comenzado la trasposición de la Directiva UE 2019/1024 sobre datos abiertos y la reutilización de la información del sector público; el Plan Estadístico Nacional 2021-2024 incluye el incremento de la explotación de los registros administrativos o el esta-

blecimiento de una arquitectura de datos comunes procedentes de distintos registros administrativos, entre otras cosas (Arroyo y Fernández Albertos, 2021; Rodríguez Póo y Salcedo Galiano, 2021). Un paso relevante en este sentido y que sienta las bases del diseño propuesto por la AIReF es el nuevo marco de colaboración establecido entre INE, Banco de España, la Agencia Estatal de la Administración Tributaria, la Seguridad Social y el Servicio Público de Empleo Estatal, que han creado un grupo de trabajo con la finalidad de que se reutilice la información estadística disponible por estos organismos por parte de los investigadores de forma que se permita el máximo aprovechamiento social dentro de lo que establece la legislación estadística.

IV. LA DIFUSIÓN Y LA UTILIZACIÓN DE LA EVALUACIÓN

En Suiza, la Ley de Libertad de Información exige que los informes de evaluación sean accesibles. Además, se dispone de una base de datos nacional (ARAMIS) que, en principio, reúne las evaluaciones, entre otra información (31). Ello no significa que las evaluaciones sean conocidas por el público general o la prensa. Aun así, las evaluaciones que solicita el Parlamento tienen una importante dimensión política ya que este puede controlar su aplicación por el Ejecutivo. La existencia de las disposiciones de evaluación y la posibilidad de solicitar una evaluación sobre determinadas medidas parece tener influencia en los debates parlamentarios, contribuyendo a obtener consensos sobre las medidas (Widmer, 2000). Numerosos estudios han tratado de estimar la utilización de las

evaluaciones llevadas a cabo por los distintos actores. Por lo que se refiere a la Administración federal, un 65 por 100 de aquellos que habían participado en una evaluación percibían que su utilización había sido elevada (Horber-Papazian y Baud-Lavigne, 2019).

En Holanda, los informes de evaluación suelen publicarse y enviarse al Parlamento (32). Los evaluadores pueden generalmente hacer uso de las evaluaciones para hacer presentaciones o escribir artículos científicos. En 2021, un Informe del Tribunal de Cuentas puso de manifiesto que, de los 26 proyectos de ley que los ministros enviaron al Parlamento durante seis meses de 2020, casi dos tercios contenían una justificación insuficiente de las medidas de política pública que proponían (Ministerio de Hacienda de Holanda, 2021). El Parlamento acostumbra a utilizar las evaluaciones. Un interesante estudio encontró que los evaluados querían que las evaluaciones de impacto recogieran también propuestas para modificar las políticas, que la evaluación sirviera para revisar la teoría del cambio (o conjunto de hipótesis en las que se apoya una intervención pública) que había dado lugar a la política y sugerían la creación de repositorios que sistematizaran los resultados de las evaluaciones (Klein Haarhuis, 2016).

En Canadá, el *Policy on Results* establece que los informes de evaluación elaborados por los ministerios y agencias deben publicarse. En Alemania, algunos organismos, como la Agencia de Cooperación Internacional, acostumbran a publicar todas sus evaluaciones, pero esto no siempre ocurre en el resto de los ministerios y el uso de las

evaluaciones varía mucho entre sectores. Los receptores de las evaluaciones son, sobre todo, los responsables de los programas evaluados y los controladores del gasto, más que los responsables políticos de la adopción de decisiones (Stockmann y Meyer, 2020). Desde la legislatura comprendida entre 2005 y 2009, el término evaluación empezó a mencionarse en las intervenciones de los parlamentarios con más frecuencia (Stockmann y Meyer, 2020). El interés del Parlamento por la evaluación ha crecido con el tiempo: las comisiones demandan con frecuencia informes científicos y pueden solicitar evaluaciones de impacto de los proyectos de ley. También puede pedir al gobierno que realice evaluación y presente sus resultados al Parlamento.

En el Reino Unido, el Servicio Gubernamental de Investigación Social (GSR) presentó un protocolo de publicaciones, actualizado en diciembre de 2021, que ha hecho que varios ministerios difundan sus evaluaciones (33). La experiencia más conocida es la desarrollada por los nueve Centros *What Works* con una estrategia clara de divulgación sistemática de sus hallazgos en los diferentes temas en los que trabajan (reducción de la criminalidad, envejecimiento, sinhogarismo, infancia, educación infantil o sanidad) (34). *What Works* ha puesto en marcha las llamadas auditorias de pruebas o de evidencia para mejorar la producción, difusión y asimilación de las evaluaciones de impacto.

Por lo que se refiere al uso de las evaluaciones, entre 2015 y 2020, la NAO manifiesta que ha realizado 967 recomendaciones, de las cuales 807 se han debatido con las organizacio-

nes evaluadas. De estas, 297 se implementaron, 473 están en proceso de implementación y 37 se rechazaron (NAO, 2022) (35). En 2016, una investigación dirigida por la Alliance for Useful Evidence reveló que la utilización de un formato simplificado, la existencia de repositorios y las estrategias de difusión y formación mejoran el uso de evidencia por los responsables públicos (36).

En Estados Unidos, los planes de evaluación de los ministerios deben incluir en la «medida de lo posible» información sobre cómo se diseminarán los resultados y se utilizarán en las decisiones sobre sus programas y políticas (OMB, 2021, p. 35). Se espera que el principio de publicar las evaluaciones se generalice en la Administración tras la Ley de Evidencia de 2018. Varios ministerios han creado centros de intercambio de información que reúnen los estudios disponibles en un campo determinado y los sintetizan en un formato accesible para no-especialistas.

Por su lado, los centros *What Works Clearinghouse* del Ministerio de Educación (37) o *CLEAR* del Ministerio de Trabajo identifican, resumen y divulgan evidencia sobre las políticas de su sector, incluyendo la proporcionada por las evaluaciones de impacto y sobre metodología de evaluación (Donaldson, Donaldson y Renger, 2022) (38).

En Francia, un análisis del impacto de las 80 evaluaciones mostró que «era posible» que hubiesen contribuido a los cambios en la política evaluada en un 44 por 100 de los casos (Delahais y Lacouette-Fougère, 2017). Aun así, según el Consejo de Estado (2020), la EvPP todavía no es una verdadera herramienta

que sirva para la adopción de decisiones.

En España, el IEPP y la AIReF publican sus evaluaciones, pero no es fácil encontrar evaluaciones en los ministerios. No existen datos sistemáticos sobre la utilización de las evaluaciones. El reciente PRTR (2021:133) del Gobierno central parece haber tomado como referencia los resultados de las evaluaciones de la AIReF en varios sectores (políticas activas de empleo, simplificación y racionalización de los incentivos a la contratación o beneficios fiscales). Además, otras evaluaciones de la AIReF han contribuido a la configuración de distintas medidas de política pública en industria, farmacia, fiscalidad, sanidad, universidades, empleo o el ingreso mínimo vital.

El PRTR ha propuesto la creación de una unidad técnica en el Ministerio de Hacienda para el seguimiento de las propuestas de la AIReF y su incorporación en el ciclo presupuestario. Se prevé también que este ministerio remita al Consejo de Ministros un informe con el estado de tales propuestas, antes del 31 de marzo de cada año. Asimismo, propone que las evaluaciones de la AIReF sean llevadas a las conferencias sectoriales, donde se reúnen los ministerios y consejerías del ramo.

Finalmente, un resumen de los principales rasgos de la EvPP de los ocho países analizados se presenta en el cuadro n.º 1.

V. DISCUSIÓN Y CONCLUSIONES

Existen buenas razones para examinar la situación de la EvPP en España en perspectiva comparada. Los últimos análisis compa-

rados ubican el caso español en un lugar modesto. Jacob, Speer y Furubo (2015) sitúan a España entre los países con un grado de madurez medio. Más recientemente, Stockmann, Meyer y Taube (2020) encuentran que España alcanza un grado medio-bajo de institucionalización de la evaluación. Sin embargo, estos estudios no incluían las iniciativas que en materia de EVPP ha puesto en marcha España en los últimos cinco años. La EvPP en nuestro país ha experimentado altibajos, pero en la actualidad atraviesa su momento de mayor apogeo en los tres últimos lustros (Casado y Del Pino, 2021). Con todo, la EvPP no es aún una práctica consolidada en España y ello hace conveniente explorar los problemas y los avances de países con más experiencia, lo cual ayudará además a calibrar los nuestros.

En este artículo se ha examinado la situación de la EvPP en ocho países. En cada uno de ellos, se han analizado distintos elementos que ayudan a valorar hasta qué punto la evaluación está institucionalizada dentro de sus sistemas político-administrativos. El artículo ha analizado el nivel central de Gobierno, aunque en algunos países como Suiza, Canadá, Alemania o España los Gobiernos subcentrales llevan crecientemente a cabo evaluaciones (Widmer, 2020; Lahey, *et al.*, 2020; Stockmann y Meyer, 2020; Casado y Del Pino, 2021, respectivamente). Otra limitación del artículo es que no se han incluido los análisis de impacto normativo que son una práctica en aumento en todos los países aquí estudiados, pero que en pocas ocasiones pueden ser considerados como EvPP (Pattyn *et al.*, 2018).

CUADRO N.º 1

INSTITUCIONALIZACIÓN DE LA EVALUACIÓN DE POLÍTICAS EN EL SISTEMA POLÍTICO-ADMINISTRATIVO DE UNA SELECCIÓN DE PAÍSES

	HOLANDA	SUIZA	CANADÁ	ALEMANIA	REINO UNIDO	ESTADOS UNIDOS	FRANCIA	ESPAÑA
Marco normativo e institucional	*Evaluación regulada en la Ley de Cuentas. *Disposiciones de EvPP en normas sectoriales. *Ministros responsables de evaluar sus políticas. *Tribunal de Cuentas: eficacia y eficiencia de las políticas. *Parlamento decide si las evaluaciones son suficientes. *Oficina de Análisis de Política Económica realiza evaluaciones.	*Constitución: Parlamento velará por la evaluación de las mediaciones de las federales. *Disposiciones de EvPP en numerosas normas sectoriales. *Parlamento: una unidad de evaluación, CPA. Ejecutivo obligado a responder. CPA coordina su evaluación con el Tribunal de Cuentas. *Ministerios realizan unas 80 evaluaciones anuales.	*Existe una política de evaluación y normativa específica. *El Consejo del Tesoro (TBS) coordina la política y los ministerios tienen estructura de evaluación. *En el Parlamento, la OAGC y la OPC realizan importantes labores de auditoría y proporcionan a los diputados análisis independientes sobre el presupuesto.	*No legislación sobre EvPP: Código Presupuestario: auditores deben valorar el logro de objetivos, impactos y rentabilidad. *Aumentan normas sectoriales con obligaciones de EvPP. *EvPP poco coordinada; ministerios: gran autonomía. *Institutos de evaluación públicos potentes en algunos sectores. *Importante papel del Tribunal de Cuentas.	*No hay regulación general sobre EvPP. *Nuevos programas incluyen evaluación. *Ministerios de Hacienda y Presidencia: dirección informal de evaluaciones. *Ministerios planifican y ejecutan sus evaluaciones. *NAO actividad de supervisión del gasto.	*Ley de Fundamentos para la Elaboración de Políticas Basadas en la Evidencia. *OMB supervisa y coordina las evaluaciones ministeriales en el Ejecutivo. *Algunos ministerios tienen potentes unidades de evaluación. *En el Parlamento, varias agencias evalúan (GAO, CBO, CRS).	*Evaluación incluida en la Constitución. *No normas concretas, pero sí disposiciones sectoriales. *Cada ministerio es responsable de sus evaluaciones. *Iniciativas concretas de EvPP, no sistemáticas. *En el Parlamento, varias agencias evalúan (GAO, CBO, CRS).	*Algunas normas mencionan la EvPP, pero no hay una norma específica: está previsto la aprobación de una en 2022. *Algunas normas sectoriales prevén EvPP. *Dos organismos de EvPP: AIREF, IEPP. *Algunos ministerios evalúan. *Tribunal de Cuentas e IGAE: auditorías y fiscalización.
Capacidad evaluadora	*Operación <i>Insight into Quality</i> para mejorar las políticas y la calidad de la evaluación. *Agenda de Evaluación Estratégica. *Estándares evaluación, guías, herramientas. *En Ministerios no hay unidades de EvPP; hay institutos de investigación semiindependientes que hacen EvPP (CPB). *También se contratan a consultoras y academia.	*2001, formalización de los estándares de evaluación. *La mayoría de las evaluaciones se contratan externamente. *Se discute si sería necesaria la acreditación de los evaluadores.	*La mayoría de las evaluaciones son internas (450 evaluadores). *TBS certifica competencia y calidad evaluadores y unidades de EvPP de ministerios. *Guías, formación, hoja de ruta para la explotación de datos.	*Evaluaciones realizadas principalmente por institutos no universitarios. *Disponibilidad de datos favorece la evaluación. *No hay falta de voluntad política o administrativa para evaluar. *No directrices generales, se están desarrollando. Existen guías.	*Intenso esfuerzo de mejora de capacidades internas, pero frecuente la subcontratación. *Tres iniciativas destacadas: <i>Policy Profession</i> (grupo de empleados públicos para el que se exige competencias en evaluación); identificar áreas de interés para investigar en los ministerios; <i>Guías Libro Verde</i>, <i>El Libro Magenta</i>.	*Agenda de Aprendizaje, cada ministerio identifica temas que necesitan evidencia. *Se evalúan capacidades ministeriales; se diseña un Plan de Evaluación Anual. *Cada ministerio: jefe de datos, responsable de evaluación y de estadística. *Trabajan para garantizar disponibilidad y calidad de datos administrativos.	*Medidas para mejorar la calidad de la evaluación. *65 por 100 de las evaluaciones se encargan a consultoras privadas. *France Stratégie elabora guías metodológicas. *CSD datos.	*Capacidad limitada para evaluar en los ministerios. *Recursos humanos limitados en las organizaciones dedicadas a evaluar. *Guías de formación, red de evaluadores. *Timidos avances disponibilidad de datos.
Uso de la evaluación	*Evaluaciones suelen publicarse y enviarse al Parlamento. *Parlamento acostumbra a utilizar las evaluaciones. *Elevada percepción de utilidad.	*Evaluaciones deben ser accesibles. Base de datos. *Parlamento puede solicitar evaluación.	*Evaluaciones deben publicarse.	*Evaluaciones no siempre se publican. *Uso desigual; los responsables de programación son los receptores, más que los políticos. *Creciente interés del Parlamento por la evaluación.	*Protocolo de publicaciones (2021). *9 <i>What Works</i>: divulgación sistemática. NAO: alto uso de las evaluaciones. *Centros <i>What Works</i> producir y diseminar evidencia.	*Planes de evaluación deben incluir información sobre difusión y uso. *Centros de intercambio de información en ministerios. *Centros <i>What Works</i> identifican, resumen y divulgan evidencia.	*Un análisis del impacto de 80 evaluaciones: impacto del 44 por 100. *Consejo de Estado cree que EvPP no es aún herramienta que sirva para las decisiones. *Previstas medidas en el PRTR que favorecerán el uso.	*No datos sistemáticos sobre publicación y uso de evaluaciones. *IEPP y AIREF publican sus evaluaciones y existe alguna evidencia de su uso. *Previstas medidas en el PRTR que favorecerán el uso.

En los países de la muestra, la EvPP se inscribe en un marco regulatorio y organizativo desigualmente estructurado y coordinado. En todos los países el ejecutivo y el legislativo realizan funciones de evaluación o que tienen que ver con ella. Canadá y Estados Unidos parecen ser los países en los que existe mayor coordinación y puede hablarse de una política de evaluación en sus Gobiernos federales. Sin embargo, no ocurre lo mismo en el resto de los países. En Alemania, Francia o el Reino Unido, las estructuras de evaluación del ejecutivo funcionan en gran medida de forma independiente.

En el caso del legislativo, la función evaluadora está más consolidada en unos países que en otros. Los tribunales de cuentas realizan funciones de auditoría y fiscalización y, en algunos países, también llevan a cabo mayor o menor número de evaluaciones. Además, en varios países, los Parlamentos tienen un papel creciente para exigir al Ejecutivo que sustente con evidencia las políticas públicas que propone.

En todos los países existe algún tipo de plasmación legal de la evaluación, pero es muy variable. En algunos, la evaluación se recoge en la constitución (Suiza y Francia), en otros existe una norma sobre el tema (Canadá y Estados Unidos). En la mayoría de los casos, la evaluación se recoge de manera más bien genérica en distintas disposiciones generales o sectoriales.

En los países donde la EvPP está más consolidada, la preocupación por si las políticas se evalúan o no ha dado paso a una creciente inquietud por la calidad y por el uso o impacto

de las evaluaciones. En el primer caso, algunos países han avanzado sustantivamente en la definición de las capacidades de los evaluadores y de las unidades de evaluación, que son examinadas y deben ser incluso acreditadas. En ocasiones, también se han definido estándares de evaluación, a veces desde las sociedades de evaluación nacionales. En todos los países existen guías y, en algunos, cajas de herramientas de evaluación, pero solo unos cuantos cuentan con unidades que apoyen sistemáticamente el desarrollo de la evaluación en los ministerios. Prácticamente en todos los países se han adoptado medidas para mejorar la disponibilidad y uso de los datos administrativos.

En cuanto a la práctica de la evaluación, en todos los países, parte de las evaluaciones que realiza el Ejecutivo se hacen internamente, mientras que otra parte se contrata con terceros en búsqueda de mayor independencia y/o calidad de la evaluación. Sin embargo, el porcentaje es variable. Además, en algunos países se contrata sobre todo con empresas privadas, pero en otros se recurre en mayor medida a institutos de investigación semipúblicos. En muchos de los Gobiernos analizados, el ejercicio de la evaluación depende en gran medida del sector de política, siendo más frecuente en algunos como empleo o cooperación al desarrollo o allí donde la disponibilidad de datos es mayor.

Por último, en todos los países el uso de la evaluación supone un importante reto. En la mayoría de ellos, hay una preocupación por la difusión, pero incluso en los países donde la publicación es obligatoria, se

reconoce que las evaluaciones no siempre están disponibles. Algunos evaluadores realizan un pormenorizado seguimiento de sus propuestas y esta información se publica. Se percibe un creciente interés por las evaluaciones en los Parlamentos nacionales que se expresa de distintos modos (más menciones, solicitudes). En algunos países, existen iniciativas sistemáticas para que todos los ministerios identifiquen los temas sobre los que requieren más evidencia y también para divulgar evidencia sólida sobre determinados programas.

NOTAS

(1) <https://www.cpb.nl/en/node>

(2) <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada.html>

(3) Un ejemplo de *DFR* puede verse aquí: <https://www.nrcan.gc.ca/transparency/reporting-and-accountability/plans-and-performance-reports/nrcan-strategic-outcomes-and-program-alignment-architecture/199>

(4) <https://www.bundesrechnungshof.de/en/audit-reports/products/annual-reports/2020-annual-report-2013-complementary-report>

(5) <https://www.bi.team/wp-content/uploads/2013/06/TLA-1906126.pdf>

(6) <https://www.bi.team/>

(7) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/753468/RiseExperimentalGovernment_Cross-GovTrialAdvicePanelUpdateReport.pdf

(8) <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>

(9) *Delivering Government Solutions in the 21st Century Reform Plan and Reorganization Recommendations*, ver página 118 en adelante <https://www.whitehouse.gov/wp-content/uploads/2018/06/Government-Reform-and-Reorg-Plan.pdf>

(10) <https://www.whitehouse.gov/omb/information-for-agencies/evidence-and-evaluation/>

(11) <https://www.dol.gov/agencies/oaspl/evaluation>

(12) <https://www.gao.gov/reports-testimonies/recommendations-database>

(13) http://sfe-asso.fr/sites/default/files/document/file/190612_sfe_barometre_2018_fichier_final_web.pdf

(14) <https://www.modernisation.gouv.fr/loffre-daccompagnement-de-la-ditp/sciences-comportementales>

(15) https://www.mptfp.gob.es/portal/ministerio/participacion_proyectos/consulta_previa/proyectos/2021/2021-05-17.html

(16) https://www.lamoncloa.gob.es/temas/fondos-recuperacion/Documents/30042021-Plan_Recuperacion_%20Transformacion_%20Resiliencia.pdf

(17) <https://www.boe.es/buscar/doc.php?id=BOE-A-2022-1639>

(18) Cemfi <https://www.cemfi.es/about/whoweare/index.asp> y J-PAL (bdul Latif Jameel Poverty Action Lab) <https://www.povertyactionlab.org/es>

(19) <https://open.overheid.nl/repository/ronl-a4047cbe-8fd5-492f-b1b0-5b07a0eaeaad/1/pdf/reflecties-en-lessen-inzicht-in-kwaliteit-2018-2021.pdf>

(20) <https://www.toolboxbeleidsevaluaties.nl/>

(21) Competencias de los evaluadores <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/evaluation-competencies.html>

(22) <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/gba-primer.html>

(23) <https://www.degeval.org/en/home/>

(24) <https://www.gov.uk/government/collections/areas-of-research-interest#full-publication-update-history>

(25) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938046/The_Green_Book_2020.pdf

(26) <https://www.gov.uk/government/publications/the-magenta-book>

(27) <https://www.gov.uk/government/publications/public-value-framework-and-supplementary-guidance>

(28) <https://oes.gsa.gov/toolkits/>

(29) <https://www.casd.eu/en/le-centre-daces-securise-aux-donnees-casd/gouvernance-et-missions/>

(30) Véase la Opinión para una estrategia de acceso a datos administrativos <https://www.airef.es/es/opiniones/estrategia-acceso-datos-administrativos/>

(31) <https://www.aramis.admin.ch/Texte/?ProjectID=28272>

(32) Véase OVERHEID.nl.

(33) <https://www.gov.uk/government/publications/government-social-research-publication-protocols>

(34) <https://www.gov.uk/guidance/what-works-network#introduction>

(35) <https://www.nao.org.uk/wp-content/uploads/2021/06/NAO-Annual-Report-and-Accounts-20-21.pdf>

(36) https://media.nesta.org.uk/documents/using_evidence_what_works.pdf

(37) <https://ies.ed.gov/ncee/wwc/FWWW>

(38) <https://clear.dol.gov/>

BIBLIOGRAFÍA

AIReF (2020). *Opinión para una estrategia de acceso a datos administrativos*. Opinión 1/20.

AIReF (2021). *La institucionalización de la evaluación de Políticas Públicas en Castilla y León: Situación actual y propuestas*. Estudio.

ALMUNIA, M. y REY BIEL, P. (2021). La gestión de los datos administrativos en España: diagnóstico y retos de futuro. *Cuadernos Económicos de ICE*, 102.

ARROYO, I. y FERNÁNDEZ ALBERTOS, J. (2021). Los retos de la explotación masiva de datos administrativos para la investigación y las políticas públicas: una visión desde dentro de la Administración. *Cuadernos Económicos de ICE*, 102.

BAÍZ, A., BONO, P.-H., DEBU, S., DESPLATZ, R., LACQUETTE-FOUGÈRE, C. y TRANNOY, A. (2019). *Vingt ans d'évaluations d'impact en France et à l'étranger. Document de travail*, 16. France Stratégie.

BLOMEYER, R. y EUSSNER, A. (2020). European Union. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp. 463-479. Cham: Palgrave Macmillan.

BANCO MUNDIAL (2019). BIN NL. <https://binnl.nl/home+-+en/default.aspx>

BOAZ, A., OLIVER, K., CUCCATO, G. y DASHWOOD, C. (2021). *Rebuilding a Resilient Britain: Data and Evaluation Areas of Research Interest across Government*. *ARI Report*, 11.

BROUSSELLE, A. y BUREGEYA, J.-M. (2018). Theory-based Evaluations: Framing the Existence of a New Theory in Evaluation and the Rise of the 5th Generation. *Evaluation*, 24(2), pp. 153-168.

BUSTELO, M. (2020). Spain. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp. 308-328. Cham: Palgrave Macmillan.

CATALÁ, D., MIGUEL DE, M. y CERVELLÓ, R. (2015). La necesidad de la evaluación de políticas públicas como herramienta de impulso a la calidad democrática. En: J. MARCO y B. NICASIO (eds.), *La Regeneración del Sistema: Reflexiones en torno a la calidad democrática, el buen gobierno y la lucha contra la corrupción*, pp. 215-232. Valencia: Asociación Valenciana de Politólogos.

CASADO, J. M. y DEL PINO, E. (2021). Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021). *Cuadernos Económicos de ICE*, 102.

CONSEJO DE ESTADO DE FRANCIA (2020.) *Conduire et partager l'évaluation des politiques publiques. Étude annuelle 2020*. París: La Documentation Française.

DELAHAIS, T. y LACQUETTE-FOUGÈRE, C. (2017). Can evaluations really contribute to evidence-based policy making at government level? The case of the French Government Modernisation Evaluations. *3rd International Conference on Public Policy, Singapore*.

DONALDSON, S. I. y RINGER, J. A. (2022). Evaluation in the United States of America. En R. STOCKMANN, W. MEYER y L. SZENTMARJAY (eds.), *The Institutionalisation of Evaluation in the Americas*, pp. 355-377. Palgrave Macmillan.

FEDEA, FUNDACIÓN DE ESTUDIOS DE ECONOMÍA APLICADA (2021). La evaluación de políticas públicas en España: antecedentes, situación actual y

- propuestas para una reforma. *FEDEA Policy Papers*, 2021/09.
- FEINSTEIN, O. y ZAPICO-GOÑI, E. (2010). Evaluation of Government Performance and Public Policies in Spain. World Bank. Independent Evaluation Group. *ECD Working Paper*, 22.
- DE FRANCESCO, F. y PATTYN, V. (2021). Policy evaluation styles. En M. HOWLETT y J. TOSUN (ed.), *The Routledge Handbook of Policy Styles*, pp.408-421. Routledge.
- GARDE-ROCA, J. A. (2017). AEVAL y su muerte anunciada. *Cinco Días*, 5 de agosto.
- GUATIER, B., LAHEY, R. E. y JACOB, S. (2022). Evaluation in Canada. En R. STOCKMANN, W. MEYER y L. SZENTMARJAY (eds.), *The Institutionalisation of Evaluation in the Americas*, pp. 143-169. Palgrave Macmillan.
- GUBA, E. G. y LINCOLN, Y. S. (1990). *Fourth Generation Evaluation*. Newbury Park: Sage.
- HERRERO, C. (2021). Gestión de recursos públicos y evaluación: un binomio inseparable. *Boletín Económico del ICE*, 3135, pp. 23-35.
- HORBER-PAPAZIAN, K. y BAUD-LAVIGNE, M. (2019). Factors Contributing to the Strong Institutionalization of Policy Evaluation in Switzerland. En A. LADNER, N. SOGUEL, Y. EMERY, S. WEERTS y S. NAHRATH (eds.), *Swiss Public Administration. Governance and Public Management*. Palgrave Macmillan.
- INSTITUTO PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS (2021). *Estudio de Diagnóstico de la Evaluación de Políticas Públicas en la Administración General del Estado*.
- JACOB, S., SPEER, S. y FURUBO, J.-E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21(1), pp. 6-31.
- JIMENO, J. F. (ed.) (2021) La evaluación de las políticas públicas en España: recursos, metodologías y algunos ejemplos. Número monográfico de *Cuadernos Económicos (ICE)*, 102. <http://www.revistasice.com/index.php/CICE/issue/view/792>
- JOHN, P. (2019). The international appeal of behavioural public policy: is nudge an Anglo-American phenomenon? *Journal of Chinese Governance*, 4(2), pp. 144-162.
- DE JONG, M. (2016). Netherlands. En D. MOYNIHAN e I. BEAZLEY (eds.), *Toward Next-Generation Performance Budgeting: Lessons from the Experiences of Seen Reforming Countries. Directions in Development*, pp. 99-110. Washington: Banco Mundial.
- KLEIN HAARHIUS, C. (2020). The Netherlands. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp. 89-114. Palgrave-Macmillan.
- LACOUETTE-FOUGÈRE, C. y SIMON, B. (2020). France. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp. 139-165. Palgrave-Macmillan.
- LAHEY, R. E. et al. (2020). Canadian Journal of Program Evaluation. *La Revue canadienne d'évaluation de programme*, 35.1, pp. 111-126.
- MINISTERIO DE HACIENDA HOLÁNDÉS (2021). Reflecties en lessen Inzicht in Kwaliteit 2018-2021. <https://open.overheid.nl/repository/ronl-a4047cbe-8fd5-492f-b1b0-5b07a0eaeed/1/pdf/reflecties-en-lossen-inzicht-in-kwaliteit-2018-2021.pdf>
- MEYER, W., STOCKMANN, R. y TAUBE, L. (2020). The Institutionalization of Evaluation. Theoretical Background, Analytical Concept and Methods. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe* (pp. 3-33). Palgrave Macmillan.
- OCDE (2021). *Evaluation frameworks and practices: a comparative analysis of five OECD countries*.
- OMB (2021) *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance* <https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf>
- ONRUBIA, J. y SÁNCHEZ-FUENTES, A.-J. (2019). Evaluación de políticas públicas: una tarea urgente e imprescindible. En A. CUENCA y S. LAGO (coords.), *El sector público español: reformas pendientes*, pp. 211-241. Madrid: Funcas.
- PATTYN, V., VAN VOORST, S., MASTENBROEK, E. y DUNLOP, C. A. (2018). Policy Evaluation in Europe. En E. ONGARO y S. VAN THIEL (eds.), *The Palgrave Handbook of Public Administration and Management in Europe*, pp. 577-598. Palgrave-Macmillan.
- DE LA RICA, S. (2020). La evaluación de las políticas públicas en el siglo XXI. *El País*, 18 de octubre.
- RUIZ, A. (2012). Panorámica actual de la evaluación de las políticas públicas. *Presupuesto y Gasto Público*, 68, pp. 13-23.
- STERN, E. (2009). Evaluation policy in the European Union and its institutions. En W. TROCHIM, M. y L. COOKSY (eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation*, 123, pp. 67-85.
- STOCKMANN, R. y MEYER, W. (2020). Germany. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp.167-198. Palgrave Macmillan.
- STOCKMANN, R., MEYER, W. y TAUBE, L. (eds.) (2020). *The Institutionalization of Evaluation in Europe*. Cham: Palgrave Macmillan.
- RODRÍGUEZ PÓO, J. M. y SALCEDO GALIANO, A. (2021). Estadística oficial y evaluación de políticas públicas: oportunidades y retos. *Cuadernos Económicos de ICE*, 102.
- VEDUNG, E. (2010). Four Waves of Evaluation Diffusion. *Evaluation*, 16(3), pp. 263-277.
- WIDMER, T. (2020). Switzerland. En R. STOCKMANN, W. MEYER y L. TAUBE (eds.), *The Institutionalization of Evaluation in Europe*, pp. 199-225. Palgrave Macmillan.
- WIRTHS, D. y HORBER-PAPAZIAN, K. (2016). Les clauses d'évaluation dans le droit des cantons suisses: leur diffusion, leur contenu et la justification à l'origine de leur adoption. <https://www.semanticscholar.org/paper/Les-clauses-d-por-100E2-por-10080-por-10099-por-100C3-por-100A9valuation-dans-le-droit-des-cantons-Wirths-Horber-Papazian/2607152509fdca4e6ac66829a63df289f03b9467>

MÉTODOS DE EVALUACIÓN DE POLÍTICAS PÚBLICAS

Joaquín ARTÉS (*)
Beatriz RODRÍGUEZ-SÁNCHEZ

Universidad Complutense de Madrid

Resumen

Las políticas públicas se diseñan para cumplir determinados objetivos. Las técnicas de evaluación de impacto causal permiten medir los efectos de las políticas públicas en las variables de interés de forma que pueda evaluarse hasta qué punto la política ha logrado sus objetivos. En este artículo se hace una revisión de los principales métodos de evaluación de impacto causal utilizados habitualmente en la investigación académica.

Palabras clave: evaluación, políticas públicas, inferencia causal.

Abstract

Public policies are implemented with the intention of achieving certain goals. The effects of such policies can be measured using various impact evaluation techniques, providing crucial information regarding the extent to which a policy has achieved such goals. In this paper, we review the main methods that are most frequently applied in academic research to evaluate the impact of public policies.

Keywords: evaluation, public policy, causal inference.

JEL classification: C50, H00.

I. INTRODUCCIÓN

LAS políticas públicas se diseñan para cumplir una serie de objetivos como pueden ser la mejora del rendimiento académico o la mejora de la salud, por citar algunos ejemplos. La evaluación de los efectos de las políticas públicas es fundamental para asegurar que la política cumple los objetivos para los que fue diseñada, para cambiar su diseño en caso de que no los cumpla y para asegurar que el dinero público se usa eficientemente.

El problema que surge a la hora de evaluar los efectos de una política pública es cómo diferenciar los cambios en la variable objetivo (e. g., el rendimiento académico) que se deben a la política implementada de aquellos cambios que se deben a otros aspectos que también pueden influir en esa variable. Un análisis que no permita analizar la relación causa-efecto entre la política pública y la variable objetivo, o lo haga erróneamente, conducirá a conclusiones incorrectas y, en consecuencia, a decisiones erróneas.

En los últimos años, la literatura académica ha desarrollado numerosos métodos de evaluación de impacto de las políticas públicas que permiten determinar con gran precisión si un cambio en una variable objetivo es causado por una determinada política pública. Pese a su importancia, el análisis riguroso de los efectos que cabe atribuir a una política pública no es habitual en el ámbito de la gestión pública en España. Con objeto de contribuir a superar esta laguna, en este artículo describimos

los principales métodos de evaluación de impacto causal que se utilizan habitualmente tanto en la literatura académica como en la práctica profesional. Comenzamos el trabajo con una descripción del problema analítico que representa la identificación de relaciones causales en el ámbito de las políticas públicas para, a continuación, hacer un repaso de los principales métodos estadísticos que permiten identificar relaciones causa-efecto.

II. EL PROBLEMA FUNDAMENTAL DE LA INFERENCIA CAUSAL EN LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

Supongamos, a efectos ilustrativos, que queremos evaluar el efecto que supone recibir clases de refuerzo durante el último año de bachillerato en el rendimiento académico obtenido en el examen de acceso a la universidad. Una simple correlación entre acudir a clases de refuerzo y el rendimiento académico obtenido en el examen podría mostrar una relación negativa entre ambas variables: a más clases de refuerzo, peor rendimiento académico en el examen de acceso. Esto podría ser debido a que los alumnos que acuden a clases de refuerzo son sustancialmente distintos a aquellos que han optado por no acudir a clases de refuerzo, dado que, previsiblemente, estos últimos son mejores estudiantes. Obviamente, sería un error interpretar esta correlación de forma causal y concluir que las clases de refuerzo empeoran el rendimiento académico.

El uso de una metodología adecuada resulta crucial en casos como el mencionado y, en general, en cualquier análisis de políticas públicas, dado que permite identificar el efecto real de la política pública (e. g., las clases de refuerzo), separando ese efecto del que pudieran tener otras variables relacionadas (e. g., la capacidad previa de los estudiantes incluidos en la comparación) en la variable objetivo.

El problema que se plantea desde el punto de vista del análisis causal es que para cada individuo que incluyamos en nuestro análisis únicamente observaremos un estado de la naturaleza: o bien el individuo ha recibido clases de refuerzo en el momento de hacer el examen de acceso, o no las ha recibido. Este problema hace que no se pueda calcular directamente el efecto causal de esa variable para cada individuo concreto. Eso no impide, sin embargo, siempre que dispongamos de una muestra suficientemente grande de individuos, que podamos calcular el efecto causal que ha tenido el programa, en media, comparando el rendimiento de un grupo de individuos beneficiario del programa (*grupo tratado*) con el de un grupo de individuos no beneficiarios del mismo que sean lo suficientemente parecidos (*grupo de control*). La razón es que al comparar dos grupos de individuos que, en media, tienen características similares en todas las variables que podrían influir en la variable objetivo, excepto el tratamiento recibido, las diferencias en la variable objetivo que se observen solo podrían ser atribuidas al tratamiento, precisamente porque, en el resto de las características, ambos grupos serían iguales.

Si ambos grupos fuesen diferentes, al comparar la nota media obtenida en el examen de acceso a la universidad de cada grupo, esta diferencia vendría explicada, en parte, por el efecto que estas otras características tienen en el rendimiento académico. En este caso, la diferencia de notas medias entre los dos grupos podría descomponerse en dos efectos: el efecto del tratamiento y el efecto de estas otras características que diferencian a ambos grupos que, siguiendo la terminología habitual, podemos denominar sesgo de selección. Es decir, la existencia de dicho sesgo hace que la diferencia de medias no coincida con el efecto causal del tratamiento que queremos analizar.

El problema fundamental de la evaluación de políticas públicas consiste, en consecuencia, en encontrar un grupo contrafactual adecuado que permita estimar de forma precisa lo que habría ocurri-

do con la variable objetivo en ausencia de la política pública que ha afectado al grupo de tratamiento, de forma que se elimine el sesgo de selección.

III. EL ESTÁNDAR DE ORO: LA ALEATORIZACIÓN

Una forma ideal de asegurar que el grupo de control es estadísticamente idéntico al grupo tratado, excepto en el tratamiento recibido, consiste en aleatorizar *ex ante* qué individuos reciben el tratamiento y cuáles no. Este es el procedimiento utilizado habitualmente en el campo de las ciencias de la salud al evaluar la eficacia de un medicamento (e. g., una vacuna). La forma habitual de medir el impacto causal de una vacuna consiste, en primer lugar, en seleccionar un grupo de población potencialmente susceptible de contraer una enfermedad. A continuación, se sortea (se aleatoriza) a qué individuos se les administra el medicamento y a cuáles no (o a cuáles se les administra un placebo). Pasado un tiempo, se comparan los porcentajes correspondientes a los individuos del grupo tratado y del grupo de control que han contraído la enfermedad contra la que la vacuna pretende ser efectiva. Una diferencia estadísticamente significativa en el porcentaje de contagios en ambos grupos puede ser interpretada de forma causal porque la aleatorización, junto con la inclusión de un número suficientemente grande de individuos en el análisis, asegura que el grupo tratado y el grupo de control son estadísticamente idénticos.

El uso de experimentos aleatorios (*RCT* o *randomized controlled trials*) en el ámbito de las ciencias sociales presenta dificultades añadidas dadas las características de los tratamientos a analizar. Sin embargo, las dificultades adicionales no han impedido que desde finales de los años noventa se haya generalizado su uso para investigar una amplia variedad de cuestiones (1). En este sentido, y a modo de ilustración, mencionaremos dos tipos de estudios representativos.

En primer lugar, la evaluación de programas piloto en el ámbito educativo o sanitario. Un programa piloto consiste en implementar una política pública a pequeña escala para poder evaluar sus efectos antes de aplicar el programa a gran escala. El programa STAR, en el estado de Tennessee, en Estados Unidos, es la referencia clásica en este tipo de estudios (véase Krueger,

1999). El programa STAR pretendía evaluar el efecto de la reducción del número de alumnos por aula en el rendimiento académico aleatorizando el número de alumnos por aula en los colegios que voluntariamente eligieron participar en el piloto. La aleatorización permitió demostrar de forma causal que la reducción del número de alumnos por aula supuso una mejora de 5 puntos porcentuales en el rendimiento académico de los alumnos que fueron asignados aleatoriamente a clases pequeñas en comparación con el grupo contrafactual. En España, sin embargo, apenas existen artículos académicos que muestren los resultados de una evaluación de políticas educativas utilizando ensayos aleatorios, aunque sí existe algún ejemplo en el ámbito de la economía de la salud. Un estudio reciente realizado en siete países europeos, entre los que se encuentra España, mostró el coste-efectividad (potenciales ahorros y mejores resultados en salud) de una intervención multimodal consistente en un programa de actividad física y consejos nutricionales en personas mayores de 70 años con limitaciones funcionales (Peña-Longobardo *et al.*, 2021).

Un segundo tipo de estudios que resulta útil mencionar, por su relevancia, son los trabajos que utilizan la aleatorización como técnica para medir la existencia de discriminación en el ámbito laboral o educativo. Una de las referencias clásicas en este contexto es el estudio de Bertrand y Mullainathan (2004). En este trabajo se pretendía medir la existencia de discriminación racial en el mercado de trabajo de Estados Unidos. El tratamiento consistió en manipular el nombre que aparecía en el *curriculum vitae* (CV) enviado a distintas ofertas de trabajo publicadas en periódicos de Boston y Chicago, siendo el resto de las características del CV idénticas entre el grupo tratado y el grupo de control. En el grupo tratado el nombre que aparecía en el CV era típicamente afroamericano, mientras que el grupo de control incluía nombres comúnmente asociados con individuos de raza blanca. El estudio muestra que las solicitudes asociadas con nombres afroamericanos recibían un 50 por 100 menos de llamadas para realizar una entrevista en comparación con solicitudes idénticas asociadas a nombres de raza blanca. En España, este tipo de aleatorización se ha utilizado, por ejemplo, para estudiar la discriminación frente al colectivo homosexual (Díaz-Serrano y Meix-Llop, 2016), o para estudiar la discriminación por razón de género (Fernández-Cornejo, 2011).

IV. EXPERIMENTOS NATURALES

En los dos tipos de análisis mencionados anteriormente se elimina el sesgo de selección mediante la aleatorización *ex ante* de tratamiento. La aleatorización *ex ante*, sin embargo, no puede llevarse a cabo para evaluar la gran mayoría de políticas públicas. Dos son las razones: o bien porque no se previó la realización de un piloto antes de la implementación de la política pública por su elevado coste o porque la naturaleza del tratamiento impide su aleatorización. En algunos casos, sin embargo, también puede llevarse a cabo una evaluación causal apoyada en la aleatorización que se produce involuntariamente, de forma «natural», por la forma en la que se llevó a cabo la política pública o por un fenómeno azaroso que permite identificar individuos que aleatoriamente han recibido o no un determinado tratamiento de interés. A este tipo de estudios se les denomina experimentos naturales.

En el ámbito de las ciencias sociales y del análisis de políticas públicas ha habido muchos ejemplos de experimentos naturales a lo largo de las últimas décadas. Uno de los ejemplos recientes más conocido es el análisis de los efectos de Medicaid, seguro público de salud estadounidense para individuos de rentas bajas (véase Finkelstein *et al.*, 2012). En el año 2008, en Oregón, se abrió de forma transitoria la posibilidad de que nuevos individuos que no tenían suscrito un seguro privado de salud pudieran beneficiarse del programa Medicaid. La respuesta fue una avalancha de nuevas solicitudes de forma que el estado de Oregón, al no poder cubrir todas las solicitudes, decidió sortear entre ellas a quienes tendrían finalmente acceso al programa. La decisión política de sortear qué individuos serían finalmente beneficiarios es equivalente a un ensayo aleatorio, aunque no se hubiese planificado previamente. Esa decisión permitió posteriormente estudiar los efectos causales de tener seguro de salud en numerosas cuestiones de interés como la frecuencia de utilización de servicios sanitarios, la salud física o mental, el empleo, el salario recibido o incluso la participación política (2).

En España, también hay circunstancias institucionales que dan lugar a aleatorizaciones naturales de interés para las ciencias sociales y las políticas públicas. Mencionaremos, a efectos ilustrativos, el análisis de los sesgos de género en la selección de los funcionarios públicos realizado por Bagues y Esteve-Volart (2010). Los distintos tribunales de evaluación de los candidatos a distintos cuerpos

funcionariales de la Administración española, desde el año 1987 al año 2007, difieren en la composición de género y, en particular, en el porcentaje de mujeres que lo forman. Los candidatos son asignados a los distintos tribunales a través de un sorteo que determina qué candidatos actuarán en primer lugar en las pruebas orales. Este sistema produce una aleatorización natural ya que los candidatos no son asignados a un determinado tribunal por su género, sino por la letra de su apellido. Este hecho hace que, *de facto*, tenga lugar una aleatorización de los candidatos a tribunales compuestos por más o menos mujeres. Bagues muestra un efecto causal de la composición de género del tribunal en la probabilidad de que candidatos de un género u otro aprueben la oposición, teniendo las mujeres menos probabilidades de aprobar, *ceteris paribus*, cuanto mayor es el porcentaje de mujeres en el tribunal.

En definitiva, el análisis causal de los efectos de determinadas políticas públicas puede analizarse de forma muy precisa y sencilla desde el punto de vista estadístico (una diferencia de medias es suficiente en la mayoría de los casos) cuando se realiza una aleatorización del tratamiento de interés entre individuos similares. Obviamente, el problema al que se enfrentan los investigadores en la mayor parte de situaciones consiste en que el grupo que recibe el tratamiento que se quiere analizar y los potenciales grupos de control no están aleatorizados, por lo que suele existir un sesgo de selección potencialmente muy severo. A continuación, analizamos una serie de técnicas estadísticas que permiten afrontar esos problemas.

V. MÉTODOS PARA CONTROLAR VARIABLES OBSERVABLES

La forma más directa de eliminar los sesgos de selección producidos por las diferencias que existen entre el grupo que recibe el tratamiento que queremos analizar y los potenciales grupos de control es utilizar un modelo estadístico que tenga en cuenta esas diferencias. Las dos técnicas que se utilizan más habitualmente para controlar por las diferencias no debidas al tratamiento entre el grupo tratado y el grupo de control son los modelos de regresión y los modelos de emparejamiento (o modelos *matching*).

1. Modelos de regresión

Los modelos de regresión siguen siendo el principal instrumento empírico con el que cuentan

los investigadores para analizar los efectos de las políticas públicas en ausencia de ensayos aleatorios o experimentos naturales. Lo usual es disponer de datos observacionales, recogidos una vez aplicada una determinada política pública en la que no se realizó ninguna aleatorización previa del tratamiento, por lo que los individuos que participan en el programa que se quiere analizar suelen ser distintos de aquellos que no participan en el programa. En estos casos, una regresión multivariante puede solventar el problema fundamental de la inferencia causal y dar lugar a estimaciones causa-efecto de las políticas públicas. Sin embargo, la inferencia a partir de una regresión únicamente será causal siempre y cuando se incluyan como controles en la regresión todas aquellas variables de control que son susceptibles de afectar a la variable de interés y que diferencian a los individuos del grupo tratado de los individuos del grupo de control. En esos casos decimos que se cumple la hipótesis de la independencia condicional.

La independencia condicional implica que cuando interpretamos una regresión mínimo-cuadrática (MCO) estamos asumiendo que, una vez que se controlan los efectos de las variables de control en la variable de interés, las diferencias en la variable objetivo entre el grupo tratado y el grupo de control se deben únicamente al tratamiento, como en un experimento aleatorizado. Podemos recurrir de nuevo al ejemplo de las clases de refuerzo para aclarar cómo podría utilizarse una regresión para evaluar efectos causales. Recordemos que en este ejemplo una diferencia de medias de la nota obtenida en el examen de selectividad por los individuos que reciben las clases de refuerzo y los que no las reciben no puede ser interpretada causalmente por la existencia de un sesgo de selección. A efectos ilustrativos, supongamos que las diferencias entre los alumnos que acuden a clases de refuerzo y los que no se deben únicamente a dos variables observables, la capacidad académica previa, y la renta de los padres (denotadas ambas por el vector X). Un modelo que tratase de medir el efecto del tratamiento (T) en el rendimiento académico (Y) podría representarse así:

$$Y = \beta_0 + \beta_1 T + \beta_2 X + e. \quad [1]$$

En este modelo, e es un error que suponemos aleatorio y los coeficientes β_0 , β_1 , y β_2 son parámetros a estimar. El coeficiente de interés es β_1 , que nos daría la diferencia entre haber recibido clases de refuerzo o no para individuos de igual renta o capa-

cidad. Es decir, el coeficiente β_1 capturaría el efecto causal del tratamiento, ya que se han controlado los efectos en la nota en el examen de selectividad que provienen de la distinta capacidad de cada estudiante y de la renta del hogar, las únicas variables que diferencian en nuestro ejemplo al grupo de control y al grupo de tratamiento.

Por el contrario, una regresión mínimo-cuadrática en la que únicamente se incluya como variable explicativa el tratamiento (haber recibido clases o no), puede representarse de la siguiente forma:

$$Y = \alpha + \beta T + e. \quad [2]$$

Comparando la ecuación [1] con la [2], es fácil deducir que la estimación por MCO de la ecuación [2] daría lugar a la siguiente relación entre el parámetro estimado (β) y el verdadero efecto del tratamiento (β_1):

$$\beta = \beta_1 + \beta_2 \text{cov}(T, X). \quad [3]$$

En este caso, β sería un estimador sesgado, que incluiría el efecto verdadero del tratamiento y el de las otras variables que diferencian al grupo tratado y al de control. El sesgo vendría producido porque el valor de la variable tratamiento depende de la renta de los padres y de la capacidad académica de cada individuo; es decir, $\text{cov}(T, X)$ no sería igual a 0.

Resulta, en consecuencia, crucial asumir que el modelo incluye todas las variables de control que diferencian al grupo tratado y al grupo de control. Obviamente, esta hipótesis es difícil de cumplir en la realidad. Esto hace de los modelos de regresión una alternativa de estimación de efectos causales netamente inferior a la aleatorización.

2. Modelos *matching*

Un modelo *matching* o de emparejamiento consiste en emparejar a cada individuo del grupo de tratamiento, no control con otro individuo (pueden ser más de uno) que no ha recibido el tratamiento, pero que sea lo más parecido en su conjunto en términos de otras características observables distintas al tratamiento (Rubin, 1973; Heckman, Ichimura y Todd, 1998). Al igual que ocurre con los modelos de regresión, para que este tipo de modelos permita identificar el efecto causal del tratamiento es necesario que se cumpla la hipótesis de independencia condicional; es decir, que la asignación al

grupo de tratamiento dependa únicamente de las variables observables que se utilizan para emparejar a los individuos. Sin embargo, al contrario de lo que sucede en los modelos de regresión, en los modelos de emparejamiento no es necesario asumir una determinada relación funcional entre las variables incluidas en el modelo, lo que reduce el riesgo de obtener resultados sesgados e inconsistentes.

Existen muchos métodos para realizar el emparejamiento entre individuos tratados y no tratados (Imbens, 2015; King y Nielsen, 2019). El método de emparejamiento más conocido es el *propensity score matching* propuesto por Rosembaum y Rubin (1983). Este método utiliza un modelo de variable dicotómica (probit o logit) para calcular la probabilidad condicional de pertenecer al grupo de tratamiento dadas las variables observadas X . A la predicción resultante de ese modelo para cada individuo se le llama *propensity score*. A partir de ese *propensity score* se asocia a cada individuo del grupo tratado con aquel individuo del grupo de control que presente el valor del *propensity score* más cercano. El objetivo final es seleccionar una muestra de individuos no tratados que sea lo más parecida posible en sus características observables a la muestra de individuos tratados, algo que puede comprobarse fácilmente a través de distintos test de diagnóstico. Uno de los requisitos fundamentales que hay que comprobar tras realizar el emparejamiento es que se cumpla la condición del *common support*: debe haber individuos suficientemente parecidos en el grupo de control para cada individuo tratado.

En el caso español existen numerosas referencias a trabajos que utilizan los métodos de emparejamiento como técnica principal de análisis. Haremos referencia, a efectos ilustrativos, a dos de estos trabajos.

El primero de estos ejemplos, en el ámbito de las energías renovables, es el trabajo de Sánchez-Braza y Pablo-Romero (2014), que analiza las diferencias en la cantidad de metros cuadrados de tierra con placas solares entre los municipios que introdujeron en el año 2010 beneficios fiscales para aquellos que las instalaran frente a los municipios que no lo consideraron. Los autores encuentran un efecto positivo de la bonificación fiscal en la instalación de placas solares, que aumentaron en más de un 70 por 100 en los municipios tratados respecto a sus pares.

Otro ejemplo en el ámbito educativo es el trabajo de García-Pérez e Hidalgo-Hidalgo (2017), que evalúa la introducción del Programa de Acompañamiento Escolar, PAE, implantado de manera progresiva desde el año 2005 hasta el año 2012, sobre el rendimiento académico. Los estudiantes tratados eran aquellos individuos de las escuelas que participaban en el PAE el mismo año en que se hacían los exámenes del programa para la evaluación internacional de estudiantes (PISA por su denominación en inglés, Programme for International Student Assessment), principalmente en el año 2011/2012. Por el contrario, el grupo control estaba compuesto de aquellos estudiantes cuyos colegios no implantaron el PAE en absoluto, pero cuyas características observables eran similares según el modelo *matching* utilizado. Los autores concluyeron que pertenecer a dicho programa de acompañamiento reportó efectos positivos en el rendimiento académico, sobre todo para aquellos que llevaban más tiempo dentro de dicho programa.

En definitiva, los métodos *matching* son una técnica muy extendida en el análisis de las relaciones causa-efecto, ya que es relativamente sencilla de aplicar y permite obtener estimaciones insesgadas siempre y cuando se tenga información suficiente, al igual que en los modelos de regresión, sobre las variables observables que determinan la pertenencia al grupo de tratamiento.

VI. MÉTODOS CUASIEXPERIMENTALES

El principal problema tanto de las técnicas de emparejamiento como de los modelos de regresión es que solo permiten garantizar una estimación insesgada del efecto causal cuando se dispone de las variables observables necesarias. En muchos casos existen diferencias no observables entre el grupo de control y el grupo tratado que impiden que las estimaciones obtenidas con modelos de regresión o modelos de emparejamiento eliminen los sesgos de selección. En efecto, supongamos que queremos estimar el siguiente modelo econométrico:

$$Y = \beta_0 + \beta_1 T + \beta_2 X + \beta_3 N + e, \quad [4]$$

donde Y , T y X tienen la misma interpretación que en la ecuación 1. Obviamente, como las características no observables, N , no pueden incluirse en el modelo, una estimación por MCO (o por métodos *matching*) que incluya únicamente las variables observables dará lugar a un sesgo de selección. En

estos casos existen técnicas estadísticas que reciben el nombre de métodos cuasiexperimentales, que permiten eliminar los sesgos de selección, incluso cuando estos sesgos se deben a diferencias no observables entre el grupo tratado y el grupo de control. A continuación, describimos tres de estas técnicas.

1. Regresión discontinua

El diseño de regresión discontinua (RD) se ha utilizado en numerosos trabajos aplicados recientes, al ser uno de los métodos cuasiexperimentales más creíbles para la identificación, estimación e inferencia de los efectos de un tratamiento (Cook, 2008; Lee y Lemieux, 2010). Este tipo de diseño fue originariamente introducido por Thistlethwaite y Campbell (1960), quienes analizaron el efecto de los reconocimientos al mérito sobre los resultados académicos futuros, asumiendo que la asignación de estos reconocimientos estuviese basada en una nota de examen o prueba de evaluación observada. La hipótesis de los autores era que los individuos con notas justo por debajo del umbral y que no recibieron el premio servían como buenos sujetos con los que comparar aquellos cuya nota estaba justo por encima y efectivamente recibieron el reconocimiento. En realidad, en el entorno cercano al umbral de corte de la variable que determina el tratamiento (variable de asignación) puede suponerse que el tratamiento está aleatorizado de forma similar a como ocurre en un ensayo aleatorio.

Esta técnica puede utilizarse para estimar efectos causales «locales». Esto significa que únicamente alrededor del umbral de corte de la variable de asignación que determina el tratamiento, las diferencias en la variable objetivo que se encuentren pueden considerarse debidas al tratamiento, ya que para individuos alejados del umbral no puede asumirse que estar a un lado u otro del umbral sea aleatorio.

Para que la inferencia obtenida a través de un diseño de regresión discontinua sea válida es necesario mostrar que se cumplen una serie de condiciones (Lee y Lemieux, 2010; Abadie y Cattaneo, 2018), como son la ausencia de manipulación en la asignación a un grupo, la ausencia de diferencias en otras variables distintas a la variable objetivo en el entorno cercano al umbral, y la consistencia de los resultados tras la realización del test de «placebo».

Aun pudiendo estimarse tanto de forma paramétrica como con métodos no paramétricos, la forma más sencilla de estimar estos modelos es a través de una regresión del siguiente tipo restringida a observaciones cuyo valor en la variable de asignación esté comprendido dentro de un determinado rango alrededor del umbral:

$$Y = \beta + \beta_1 T + \beta_2 \text{Variable de asignación} + e. \quad [5]$$

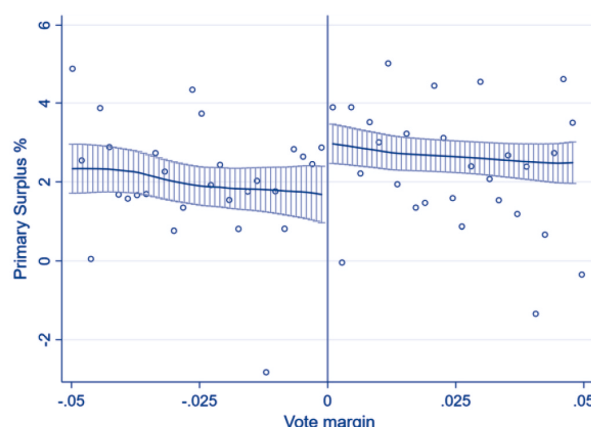
En esta regresión, al incluir como regresor la variable de asignación, el efecto del tratamiento, β_1 , se estaría estimando de forma local; es decir, justo en el umbral que determina la pertenencia al grupo de tratamiento.

Si bien la estimación por MCO es muy sencilla, la forma más habitual de estimar un modelo de regresión discontinua es de forma no paramétrica; y los resultados se suelen representar gráficamente. Como ilustración, en el gráfico 1 mostramos los resultados de estimar un modelo no paramétrico de regresión discontinua extraído del trabajo de Artés y Jurado (2018). Este trabajo analiza el efecto en el gasto público y en los ingresos públicos de la existencia de Gobiernos de coalición frente a Gobiernos de un único partido en municipios españoles. La variable de «asignación» utilizada en este trabajo fue el porcentaje de votos que el partido ganador obtuvo por encima o por debajo del porcentaje mínimo de votos necesario para gobernar en solitario. El gráfico 1 ilustra el efecto causal de pasar de tener un Gobierno en coalición (lado izquierdo del umbral) a gobernar en solitario (lado derecho de la línea roja) sobre el excedente presupuestario (diferencia entre ingresos y gastos públicos). Puede observarse que justo en el umbral que determina el tratamiento (pasar de Gobierno de coalición a mayoría absoluta) se produce una discontinuidad o salto en el excedente presupuestario. La cuantificación de la magnitud y la significatividad del efecto en un modelo de regresión discontinua sería la diferencia entre la línea ajustada a un lado y a otro del umbral, pudiendo estimarse con un modelo no paramétrico. Una buena descripción del método puede encontrarse en Cattaneo *et al.* (2016).

Finalmente, concluimos este breve repaso a los diseños de regresión discontinua haciendo referencia a otro trabajo reciente que ha utilizado esta técnica en el caso español. Se trata del trabajo de Curto-Grau *et al.* (2018) que analiza las transferencias de gasto entre Gobiernos regionales y municipales. El trabajo muestra que los munici-

GRÁFICO 1

REPRESENTACIÓN GRÁFICA DE UN ESTUDIO EN EL QUE SE APLICA REGRESIÓN DISCONTINUA



Nota: El eje horizontal muestra el margen de votos respecto al número de votos para alcanzar la mayoría y gobernar en solitario, mientras que el eje vertical representa el excedente presupuestario.
Fuente: Artés y Jurado (2018).

pios ideológicamente alineados con el Gobierno regional obtienen transferencias por parte de ese Gobierno que son hasta un 100 por 100 superiores. Para identificar el efecto causal, los autores estudian únicamente Gobiernos municipales en los que el bloque ideológico que finalmente acabó gobernando se determinó por muy pocos votos, siendo la variable de asignación el número de votos que habrían hecho falta para que cambiase el bloque ideológico que controla el municipio. Los autores también muestran que el efecto es mucho mayor cuando el Gobierno regional gobierna con una mayoría holgada, llegando a desaparecer cuando el Gobierno regional obtuvo la mayoría por muy poco margen.

2. Diferencias en diferencias

El método de diferencias en diferencias (*DiD*, por su denominación en inglés, *differences-in-differences*) es otra alternativa enormemente popular para evaluar el impacto causal de una política pública incluso en presencia de factores no observados. Esa metodología resulta especialmente adecuada cuando el tratamiento se aplica a un grupo de individuos, pero no a otros a partir de un determinado momento del tiempo (Heij *et al.*, 2004; Rabe-Hesketh y Skrondal, 2008). Normalmente, estas

políticas se aplican al grupo de tratamiento a partir de un momento del tiempo, por lo que se pueden observar las diferencias en la variable objetivo entre el grupo tratado y el grupo de no tratados antes de la implementación de la política pública, asumiendo que dichas diferencias debidas a variables no observables cuando no hay tratamiento son constantes a lo largo del tiempo.

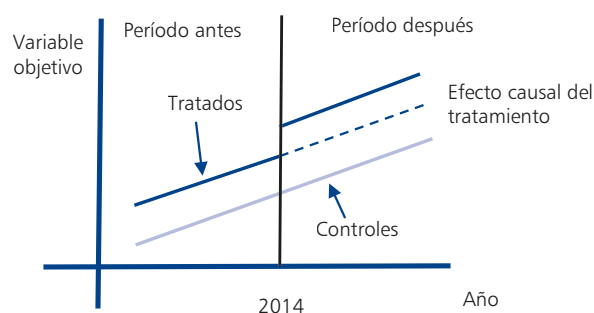
En términos de regresión, el modelo de diferencias en diferencias puede expresarse de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 \text{Después} + \beta_3 T_i * \text{Después}_i + \beta_4 X'_i + \varepsilon_i. \quad [6]$$

Y_i representa la variable objetivo de que se trate. El vector X incluye variables de control que permiten controlar por diferencias observables entre cada individuo. T es una variable dicotómica que tiene valor 1 para los individuos que pertenecen al grupo de tratamiento y 0 para los que no. Nótese que la variable T tiene el mismo valor en el período antes del tratamiento y en el período posterior; es decir, no mide el efecto del tratamiento, sino que controla por las diferencias no observables que existen de forma permanente entre los individuos del grupo tratado y los demás. La variable *Después* es otra *dummy* que adopta valor 1 para los períodos posteriores a la implementación del tratamiento y 0 para los anteriores. Esta variable captura los *shocks* temporales que pudieran afectar a la variable objetivo a lo largo del tiempo y que son comunes tanto al grupo tratado como al grupo de control. La interacción de T y *Después* identifica a las observaciones del grupo tratado durante los períodos de tratamiento. β_3 es, en consecuencia, el coeficiente de interés, que permite aislar el efecto causal.

Al igual que ocurre con los diseños de regresión discontinua, para que la inferencia obtenida a través de un diseño de diferencias en diferencias sea válida es necesario mostrar que se cumplan una serie de condiciones. La más importante es la existencia de tendencias paralelas entre el grupo tratado y el grupo de control. Ambos grupos no tienen por qué ser iguales, de ahí la gran ventaja de este estimador, pero debe mostrarse que las diferencias en la evolución de la variable objetivo en el período pretratamiento son constantes a lo largo del tiempo. Existen varios test econométricos que permiten comprobar que se cumple esta condición, pero en muchas ocasiones es suficiente con un gráfico que

GRÁFICO 2
REPRESENTACIÓN GRÁFICA DEL MODELO
DE DIFERENCIAS EN DIFERENCIAS



Fuente: Elaboración propia.

muestre la evolución de la variable objetivo. En el gráfico 2 se ofrece una representación gráfica del modelo de diferencias en diferencias.

Supongamos una política pública que se aplica a un grupo tratado en el año 2014. Idealmente, el grupo de control (la línea azul claro) y el grupo tratado (línea azul oscuro) deben presentar una evolución similar en la variable objetivo antes del tratamiento. Si esto es así, el grupo de control sería un contrafactual adecuado del grupo tratado una vez que se tengan en cuenta las diferencias constantes entre ambos grupos a lo largo de todo el período.

En el contexto español existen numerosos trabajos que han utilizado la técnica de diferencias en diferencias para evaluar políticas públicas de los que, a modo de ejemplo, destacamos dos. En el reciente trabajo de García-Vega, Kneller y Stiebale (2021), los autores tenían como objetivo evaluar el efecto del contrato de emprendedores contemplado en la reforma laboral española del año 2012 en la innovación. Los autores concluyeron que el número de innovaciones de producto aumentó de manera significativa para las empresas tratadas (aquellas que se podían beneficiar de dicho contrato) en el período posterior a la reforma laboral introducida en España en el año 2012 respecto al grupo control.

Otro ejemplo es el análisis sobre el impuesto a las bebidas azucaradas implantado en marzo del año 2017 en Cataluña (Fichera *et al.*, 2021). Usando información sobre el consumo de 26 tipos de bebidas, tanto azucaradas como no azucaradas, los autores encuentran una reducción en el con-

sumo de bebidas azucaradas tras la introducción del tributo sobre este tipo de bebidas, mostrando diferencias según el tipo de bebida.

3. Variables instrumentales

Una tercera técnica que permite en principio estimar el efecto causal de un tratamiento en una variable de interés, incluso cuando existen variables no observables omitidas en el modelo, son los modelos de variable instrumental. Estos modelos se apoyan en la siguiente lógica: para comprobar si los cambios en una variable causan los cambios en la otra, hay que identificar una tercera variable (el instrumento) que afecte a la probabilidad de recibir el tratamiento, pero no a otras variables que puedan afectar a la variable de interés (Martens *et al.*, 2006; Imbens, 2014). La razón por la que esta variable debe afectar a la probabilidad de recibir el tratamiento, pero no a la variable de interés, es porque queremos comprobar, precisamente, si los cambios en el tratamiento provocados por cambios en esta tercera variable afectan a la variable de interés únicamente a través del efecto que ha producido el instrumento en el tratamiento. Si el instrumento afecta directamente tanto la variable de tratamiento como a la variable objetivo, el instrumento es inválido, ya que no permite distinguir si los cambios en la variable objetivo se deben a los cambios en el tratamiento o al efecto directo en la variable objetivo de esa tercera variable. Si el instrumento solo afecta directamente al tratamiento, entonces sí observamos un cambio en la variable objetivo cuando cambia el instrumento y, por tanto, podemos estar seguros de que la causa del cambio en la variable objetivo es precisamente el tratamiento.

Una vez aclarado intuitivamente en qué consiste un modelo de variable instrumental, debe aclararse también que una variable solo puede servir como instrumento cuando presenta dos características estadísticas: en primer lugar, la validez (ha de afectar solo al tratamiento, pero no directamente a la variable objetivo) y, en segundo lugar, la fortaleza (debe afectar lo suficiente a la probabilidad de recibir el tratamiento como para poder extraer conclusiones estadísticamente relevantes).

Un ejemplo clásico del uso de variable instrumental es el análisis de Angrist y Kruger (1991) del efecto de los años de escolarización en el salario de los individuos. Angrist y Krueger observaron que la fecha de nacimiento de cada niño determina exó-

genamente los años de escolarización. En efecto, pensemos en el caso de España: un niño nacido en enero recibirá, en media, un año más de escolarización obligatoria que un niño nacido en diciembre. Los autores argumentan que como la fecha de nacimiento no tiene efectos en la habilidad o capacidad innata del individuo, característica no observable que codetermina ambas variables, puede utilizarse como instrumento ya que cumpliría las condiciones de validez y fortaleza. El trabajo muestra, utilizando datos de EE. UU., que los años adicionales de escolarización tienen un efecto positivo y significativo en el salario posterior de los individuos.

Un ejemplo reciente de utilización de variable instrumental para el caso español es Artés (2014). En este trabajo se trata de analizar a qué partido político favorece la abstención estimando un modelo en el que la variable dependiente es el porcentaje de voto a distintos partidos en cada municipio en distintas elecciones generales. La variable de interés o de tratamiento es el porcentaje de abstención. El instrumento utilizado es la meteorología; en concreto, el porcentaje de lluvia en cada municipio en el día de las elecciones. Este instrumento afecta a la abstención (los días de lluvia es más costoso ir a votar, por lo que existe una relación clara entre la lluvia y el porcentaje de participación electoral), pero no al sentido del voto (las personas que acuden a votar no deciden a quién votar en función de si llueve o no ese día). El cumplimiento de la condición de fortaleza y validez permite estimar causalmente el efecto de la abstención en el porcentaje de voto a cada partido. El autor encuentra que la abstención, en general, beneficia a los partidos de derechas, y perjudica, en mayor medida, a los partidos no tradicionales.

VII. DESARROLLOS RECIENTES

En este trabajo se ha mostrado no solo que las técnicas para evaluar los posibles efectos de las políticas públicas son numerosas, sino que experimentan una constante evolución y desarrollo. En línea con algunos de los métodos antes descritos, en esta sección se van a mencionar brevemente algunos desarrollos metodológicos recientes en este ámbito, que están en la mayoría de los casos relacionados de una manera u otra con las técnicas antes descritas.

En el ámbito de los modelos *matching*, existen una variedad de desarrollos recientes que tratan de mejorar algunas limitaciones de los modelos tradicionales. Una de estas técnicas es el emparejamiento

genético o *genetic matching* que corrige algunas limitaciones de los modelos que utilizan el *propensity score* (Diamond y Sekhon, 2013). En primer lugar, dicha técnica elimina la necesidad de verificar de forma manual e iterativa el *propensity score*, ya que, por un lado, proporciona un algoritmo de búsqueda automatizado que identifica las mejores coincidencias y maximiza el equilibrio en las variables observables entre el grupo tratado y el grupo de control. Esto supone estimaciones menos sesgadas (Radice et al., 2011; Diamond y Sekhon, 2013).

Siguiendo una lógica similar a los modelos de diferencias en diferencias, el método del control sintético ha sido descrito como «el desarrollo más importante en la evaluación de programas en la última década» (Athey e Imbens, 2017). Esta técnica, a grandes rasgos, consiste en construir un grupo de control «sintético» asignando distintos pesos a las observaciones no tratadas (a diferencia de modelo de diferencias en diferencias en el que cada observación tiene el mismo peso) de forma que en su conjunto se consiga un contrafactual lo más similar posible al grupo tratado. Una descripción de las ventajas de este método y de sus aplicaciones puede encontrarse en Abadie (2021). Otro desarrollo reciente en el ámbito de los estimadores de diferencias en diferencias son las mejoras en la estimación de modelos en los que el tratamiento se aplica a distintas unidades del grupo tratado en distintos momentos del tiempo. En estos casos la hipótesis de que las tendencias son paralelas no siempre permite identificar el efecto causal del tratamiento, por lo que se han desarrollado métodos que permiten superar estos problemas como los desarrollados por Sant'Anna y Zhao (2020), Callaway y Sant'Anna (2021) o Athey e Imbens. (2022)

Finalmente, cabe mencionar la utilidad para los análisis causa-efecto de desarrollos recientes en el ámbito de los modelos de *big data*. Las técnicas de *machine learning* consisten en utilizar una submuestra, llamada de «entrenamiento» o *training sample*, para predecir el comportamiento de una variable de interés. Una vez que se ha validado la capacidad predictiva en la submuestra, puede aplicarse ese modelo para predecir la evolución de la variable objetivo a gran escala. Aunque estas técnicas tienen como objetivo la predicción y no la identificación de efectos causales, tienen un potencial enorme para ayudar a construir grupos de control enormemente parecidos al grupo tratado que se quiere analizar, permitiendo que el contrafactual estimado se acerque al ideal de la aleatorización en

mayor medida. Una revisión de estos métodos y de su potencial para mejorar la evaluación de políticas públicas puede encontrarse en Athey e Imbens (2017: pp. 22-27).

VIII. CONCLUSIONES

La evaluación del impacto causal de las políticas públicas es fundamental para mejorar su implementación. En este trabajo se ha hecho una revisión de los principales métodos de estimación de relaciones causa-efecto en este ámbito. Todos estos métodos pretenden encontrar un contrafactual adecuado que permita estimar qué habría ocurrido con la variable de interés si no se hubiera llevado a cabo la política pública analizada.

La aleatorización *ex ante* del grupo tratado y del grupo de control es el mejor método para encontrar un contrafactual adecuado. Sin embargo, como la aleatorización no es posible en la gran mayoría de casos, se han desarrollado técnicas que bajo determinadas condiciones permiten la identificación precisa de los efectos causales de una política pública, incluso cuando existen diferencias no observables entre el grupo de individuos beneficiario de la política pública y los no beneficiarios.

En la actualidad, la utilización de estos métodos en España no es muy frecuente más allá del ámbito académico. Sería deseable su aplicación en el ámbito de la gestión pública, tanto para mejorar el diseño como la aplicación de las políticas públicas. Un conocimiento adecuado de los efectos de los programas públicos redundaría en una mejora de la toma de decisiones por parte de los gestores políticos y en consecuencia del bienestar general.

NOTAS

(*) JOAQUÍN ARTÉS agradece la financiación recibida a través del proyecto CSO2017-82881_R.

(1) En el año 2019 se le concedió el premio Nobel de Economía a ESTHER DUFO, ABHIJIT BANERJEE y MICHAEL KREMER por el uso de ensayos aleatorizados (*randomized controlled trials*, o *RCT*, por sus siglas en inglés) en el ámbito de las políticas públicas.

(2) Una descripción del programa y una lista completa de publicaciones relacionadas puede consultarse en la página del NBER dedicada a este experimento: <https://www.nber.org/programs-projects/projects-and-centers/oregon-health-insurance-experiment?page=1&perPage=50>

BIBLIOGRAFÍA

ABADIE, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), pp. 391-425.

- ABADIE, A. y CATTANEO, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10, pp. 465-503.
- ALBERT, R., ESCOT, L. y FERNÁNDEZ-CORNEJO, J. A. (2011). A field experiment to study sex and age discrimination in the Madrid labour market. *The International Journal of Human Resource Management*, 22(2), pp. 351-375.
- ANGRIST, J. D. y KEUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), pp. 979-1014.
- ARTÉS, J. (2014). The rain in Spain: Turnout and partisan voting in Spanish elections. *European Journal of Political Economy*, 34, pp. 126-141.
- ARTÉS, J. y JURADO, I. (2018). Government fragmentation and fiscal deficits: a regression discontinuity approach. *Public Choice*, 175(3), pp. 367-391.
- ATHEY, S. e IMBENS, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), pp. 3-32.
- ATHEY, S. e IMBENS, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1), pp. 62-79.
- BAGUES, M. F. y ESTEVE-VOLART, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies*, 77(4), pp. 1301-1328.
- BERTRAND, M. y MULLAINATHAN, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), pp. 991-1013.
- CALLAWAY, B. y SANT'ANNA, P. HC. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225, pp. 200-300.
- CATTANEO, M. D., TITIUNIK, R., VÁZQUEZ-BARE, G. y KEELE, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, 78(4), pp. 1229-1248.
- COOK, T. D. (2008). «Waiting for life to arrive»: a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2), pp. 636-654.
- CURTO-GRAU, M., SOLÉ-OLLÉ, A. y SORRIBAS-NAVARRO, P. (2018). Does electoral competition curb party favoritism? *American Economic Journal: Applied Economics*, 10(4), pp. 378-407.
- DIAMOND, A. y SEKHON, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3), pp. 932-945.
- DÍAZ-SERRANO, L. y MEIX-LLOP, E. (2016). Do schools discriminate against homosexual parents? Evidence from a randomized correspondence experiment. *Economics of Education Review*, 53, pp. 133-142.
- FICHERA, E., MORA, T., LÓPEZ-VALCÁRCEL, B. G. y ROCHE, D. (2021). How do consumers respond to «sin taxes»? New evidence from a tax on sugary drinks. *Social Science & Medicine*, 274, 113799.
- FINKELSTEIN, A., TAUBMAN, S., WRIGHT, B., BERNSTEIN, M., GRUBER, J., NEWHOUSE, J. P., ALLEN, H., BAICKER, K. y OREGON HEALTH STUDY GROUP (2012). The Oregon health insurance experiment: evidence from the first year. *The Quarterly Journal of Economics*, 127(3), pp. 1057-1106.
- GARCÍA-PÉREZ, J. I. e HIDALGO-HIDALGO, M. (2017). No student left behind? Evidence from the Programme for School Guidance in Spain. *Economics of Education Review*, 60, pp. 97-111.
- GARCÍA-VEGA, M., KNELLER, R. y STIEBALE, J. (2021). Labor Market reform and innovation: Evidence from Spain. *Research Policy*, 50(5), 104213.
- HECKMAN, J. J., ICHIMURA, H. y TODD, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2), pp. 261-294.
- HEIJ, C., DE BOER, P., FRANCES, P. H., KLOEK, T. y VAN DIJK, H. K. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.
- IMBENS, G. W. (2014). Instrumental variables: an econometrician's perspective (n.º w19983). *National Bureau of Economic Research*.
- IMBENS, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, 50(2), pp. 373-419.
- KING, G. y NIELSEN, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27(4), pp. 435-454.
- KRUEGER, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2), pp. 497-532.
- LEE, D. S. y LEMIEUX, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), pp. 281-355.
- MARTENS, E. P., PESTMAN, W. R., DE BOER, A., BELTSE, S. V. y KLUNGEL, O. H. (2006). Instrumental variables: application and limitations. *Epidemiology*, 260-267.
- PEÑA-LONGBARDO, L. M., OLIVA-MORENO, J., ZOZAYA, N., ARANDA-RENEO, I., TRAPERO-BERTRAN, M., LAOSA, O., SINCLAIR, A. y RODRÍGUEZ-MAÑAS, L. (2021). Economic evaluation of a multimodal intervention in pre-frail and frail older people

with diabetes mellitus: the MID-FRAIL project. *Expert Review of Pharmacoeconomics & Outcomes Research*, 21(1), pp. 111-118.

RABE-HESKETH, S. y SKRONDAL, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

RADICE, R., RAMSAHAI, R., GRIEVE, R., KREIF, N., SADIQUE, Z., y SEKHON, J. S. (2012). Evaluating treatment effectiveness in patient subgroups: a comparison of propensity score methods with an automated matching approach. *The International Journal of Biostatistics*, 8(1), p. 25.

ROSENBAUM, P. R. y RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp. 41-55.

RUBIN, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1), pp. 185-203.

SÁNCHEZ-BRAZA, A. y PABLO-ROMERO, M. D. P. (2014). Evaluation of property tax bonus to promote solar thermal systems in Andalusia (Spain). *Energy Policy*, 67, pp. 832-843.

SANT'ANNA, P. H. y ZHAO, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), pp. 101-122.

THISTLETHWAITE, D. L. y CAMPBELL, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), p. 309.

Resumen

El presente artículo realiza un análisis de la evidencia reciente sobre las consecuencias no intencionadas de las políticas sanitarias y educativas. Mediante varios ejemplos reales, examinamos algunos de los mecanismos comunes que provocan efectos inesperados y potencialmente indeseables. Entre los mecanismos analizados se encuentran las respuestas conductuales heterogéneas ante las mismas políticas públicas, las externalidades y los efectos de arrastre, la interacción entre distintas políticas, y la incorrecta generalización de los resultados de determinadas investigaciones. Nuestro trabajo pone de manifiesto la necesidad de considerar múltiples mecanismos y resultados al evaluar el efecto causal de las políticas públicas.

Palabras clave: consecuencias no intencionadas, análisis de políticas públicas, educación, sanidad.

Abstract

This paper reviews recent evidence on the unintended consequences of health and education policies. Through a series of topical examples, we explore common mechanisms driving unexpected and potentially undesirable effects. The mechanisms we discuss include: heterogenous behavioural responses; externalities and spillovers; policy interactions; and the incorrect generalization of research findings. Our exploration suggests the need to account for a multiplicity of mechanisms and outcomes when evaluating the causal impact of policies.

Keywords: unintended consequences, public policy analysis, education, health.

JEL classification: H50, I18, I28.

CONSECUENCIAS NO INTENCIONADAS DE LAS MEDIDAS DE POLÍTICA PÚBLICA: EJEMPLOS DE SANIDAD Y EDUCACIÓN

Jennifer GRAVES (*)

Universidad Autónoma de Madrid (UAM)

Eugenio ZUCHELLI

Madrid Institute for Advanced Study (MIAS),

Universidad Autónoma de Madrid (UAM), University of Lancaster, IZA

I. LA IMPORTANCIA DE TOMAR DECISIONES BASADAS EN LA EVIDENCIA

DETERMINAR correctamente los efectos causales de las políticas públicas resulta fundamental para comprender su capacidad para conseguir los objetivos que constituyen la esencia motivadora de dichas políticas. Por ejemplo, cuando un país aplica una política sanitaria resulta fundamental conocer si tiene un impacto real en la salud, o cuando aplica una política educativa resulta importante determinar si ha mejorado el aprendizaje del alumnado según lo previsto. En este artículo revisamos algunos ejemplos provenientes de la literatura académica sobre políticas públicas en materia de sanidad y educación para mostrar la importancia de determinar e interpretar correctamente los efectos causales de las políticas públicas.

Aunque muchas de estas políticas funcionan según lo esperado, en cuyo caso la cuestión pertinente a la hora de formular políticas es si los beneficios justifican los costes de tales programas, en ocasiones las medidas de política pública pueden aca-

rear consecuencias no intencionadas. Cuando se dan estos casos, la política en cuestión puede tener un impacto contrario al que esperaban los responsables políticos, o quizá alcance el objetivo marcado, pero también cause «efectos secundarios» imprevistos que deben tenerse en cuenta en el cálculo de la relación coste-beneficio de la medida. En este trabajo proporcionamos ejemplos de cómo la mejor de las voluntades puede acabar dando lugar a un mal resultado en la práctica y de cómo tanto investigadores como responsables políticos pueden aprender de los errores pasados para evitar problemas comunes que den lugar a consecuencias no intencionadas en la aplicación de políticas.

En la primera parte del trabajo, revisamos una serie de políticas sanitarias que han tenido importantes consecuencias no intencionadas debido a que la forma en la que se pusieron en práctica no interpretaba correctamente la evidencia científica. En concreto, mostramos la importancia que tiene en la implementación de políticas públicas tener en cuenta los siguientes aspectos: i) la heterogeneidad de tratamiento; ii) los efectos arrastre y las potenciales exter-

nalidades; y iii) la interacción entre regulaciones en distintos niveles de Gobierno (por ejemplo, autonómico/local vs. nacional). En la segunda parte del trabajo, revisamos distintos ejemplos de políticas educativas para mostrar la importancia de tener en cuenta los siguientes aspectos: i) considerar el abanico de respuestas conductuales que se derivarán de la política pública; ii) comprender adecuadamente cómo se generalizarán los resultados en otros contextos distintos a los específicamente investigados en los trabajos de investigación; y iii) tener en cuenta la multiplicidad de objetivos que puede cumplir una política educativa, y en particular cómo la educación de los niños pequeños también hace las veces de servicio de cuidado infantil durante la jornada laboral y cómo esto puede afectar a las decisiones familiares.

Los ejemplos proporcionados en las siguientes secciones (sección segunda sobre políticas sanitarias y sección tercera sobre políticas educativas) se escogieron con el objeto de mostrar que no tener en cuenta determinados factores a la hora de determinar los efectos causales de una política pública implicaría una incorrecta interpretación de la repercusión total real de esa política y podría conducir a tomar decisiones inadecuadas. Los estudios académicos seleccionados no pretenden realizar una revisión exhaustiva de la literatura académica sobre políticas sanitarias y educativas, sino ilustrar con determinados ejemplos algunos aspectos comunes que deberían tenerse en cuenta en todas las medidas de política pública. De hecho, cualquiera de los mecanismos analizados mediante el uso de políticas sa-

nitarias podría haberse expuesto de igual manera usando medidas de política educativa, y viceversa. Nuestro objetivo con el presente estudio es centrar la atención en lo habituales y relevantes que son estas cuestiones en la aplicación de cualquier política pública en general.

II. EJEMPLOS DE POLÍTICAS SANITARIAS

Pese al rápido aumento de la bibliografía acerca de la evaluación del impacto de las políticas sanitarias, sigue habiendo desafíos metodológicos y prácticos que deberían tenerse en cuenta para poder interpretar correctamente la evidencia disponible sobre el impacto causal de las políticas sanitarias. Como hemos mencionado anteriormente, en esta sección nos centraremos en la importancia de considerar: i) la heterogeneidad de respuestas conductuales ante una misma política pública; ii) las potenciales externalidades y efectos de arrastre; y iii) la interacción entre leyes a distintos niveles de Gobierno (por ejemplo autonómico/local vs. nacional). Para ello, analizamos las consecuencias no intencionadas de varias medidas de salud pública de gran alcance, incluidas las destinadas a reducir el daño provocado por el consumo de alcohol y tabaco, así como determinadas políticas de seguridad pública.

1. Efectos heterogéneos del tratamiento

Adda y Cornaglia (2010), posiblemente, uno de los estudios más citados acerca de las respuestas conductuales no intencionadas y potencialmente perniciosas de las políticas anti-tabaco. Los autores analizan los

efectos de las prohibiciones de fumar y de los impuestos especiales en la exposición al humo del tabaco de los *no fumadores* en Estados Unidos. Su principal conclusión es que la prohibición de fumar en espacios públicos (lugares de trabajo, bares y restaurantes) desplazó a los fumadores de los espacios públicos a los privados (sus hogares) y en realidad *incrementó la exposición al humo de los fumadores pasivos*, incluidos menores. Cabe destacar que los autores examinan específicamente la *heterogeneidad* de los efectos de las políticas antitabaco y que los resultados se presentan en base al consumo de tabaco en el hogar (familias fumadoras vs. no fumadoras), grupo de edad (menores vs. adultos) y día de la semana (días laborables vs. fines de semana).

Entre los diversos hallazgos interesantes presentados en el estudio, los autores constatan que *ninguna* de las dos políticas antitabaco –prohibiciones de fumar e impuestos especiales– parecen tener un impacto en las personas que viven en hogares no fumadores si se consideran conjuntamente todos los grupos de edad. Sin embargo, esta fusión de los grupos de edad enmascara una importante heterogeneidad en los efectos. En concreto, también observan aumentos estadísticamente significativos en los niveles de cotinina: i) durante los días laborables, en los menores que conviven con fumadores tras la implantación de prohibiciones de fumar más estrictas en lugares de trabajo; y ii) durante los fines de semana, en los menores que viven en familias no fumadoras tras la implantación de prohibiciones en bares y restaurantes.

Estos resultados evidencian el desplazamiento de los fumadores desde los espacios públicos a los domicilios particulares. Los autores apuntan que, tras la implantación de estas prohibiciones, los fumadores tienden a pasar menos tiempo fumando en lugares de trabajo, bares y restaurantes y más tiempo en domicilios particulares, lo cual incrementa la exposición pasiva al tabaco de sus propios hijos y de los hijos de las familias no fumadoras. A la vista del nivel de cotinina detectado en los menores, este efecto desplazamiento supone un aumento de alrededor de 2-4 cigarrillos diarios fumados en su presencia. A su vez, a diferencia de los resultados que se obtienen al analizar las prohibiciones de fumar, este mismo estudio concluye que la subida de los impuestos del tabaco reduce la exposición media al humo de los fumadores pasivos. En consecuencia, los resultados constatan efectos contrarios a los previstos en menores y *efectos potencialmente contrapuestos* de dos políticas públicas que tienen el mismo objetivo inicial.

Un estudio posterior de Carpenter, Postolek y Warman (2011) emplea un enfoque de diferencias en diferencias similar al utilizado en Adda y Cornaglia (2010), para analizar la variación temporal y geográfica provocada por la implantación de las prohibiciones de fumar en Canadá. Este estudio obtiene un resultado distinto, ya que no encuentra evidencias de efecto desplazamiento del humo ambiental del tabaco hacia los hogares. El estudio, sin embargo, se apoyaba en datos de encuesta reportados por los individuos y no incluía marcadores biológicos como la cotinina. Por este motivo, el uso de la concentración de cotinina

en fluidos corporales (un biomarcador usado para medir la exposición al humo ambiental) utilizado por Adda y Cornaglia (2010) resulta más fiable desde el punto de vista científico ya que los efectos nocivos en los menores del hogar se identificaron mediante los aumentos estadísticamente significativos de los niveles de cotinina. Dichos aumentos aparecían de forma clara en el análisis estadístico, pero es posible que fueran imperceptibles para los padres y, en consecuencia, no serían recogidos en los datos de encuesta.

Cabe señalar que Adda y Cornaglia pretendían evaluar los efectos de las políticas antitabaco en no fumadores, ya que en el momento de la publicación de su estudio no había pruebas sólidas sobre la cuestión, especialmente en la literatura económica. Desde entonces, la literatura económica sobre los efectos causales, tanto previstos como no previstos, de las políticas antitabaco ha crecido de manera notable y ahora ya existen numerosos estudios sobre esta cuestión. Hoy en día se reconoce ampliamente que las prohibiciones de fumar en espacios públicos parece que únicamente reducen el consumo de tabaco en determinados subgrupos, como es el caso de las personas que acuden con frecuencia a bares y restaurantes o que fuman mucho (p. ej., Anger, Kvasnicka y Siedler, 2011; De Chaisemartin, Geoffard y Le Faou, 2011; Jones *et al.*, 2015).

Los resultados de la literatura sobre las políticas antitabaco son un ejemplo ilustrativo, pero existen otros ejemplos de política pública en materia de salud en los que también se han encontrado respuestas conductuales no previstas. Por ejemplo,

Carpenter y Stehr (2011) utilizan un diseño cuasiexperimental basado en los distintos momentos en los que se aprobaron las leyes que introducían la obligatoriedad del uso del casco para montar en bicicleta en los distintos estados de Estados Unidos. Los autores se centran en los efectos de estas políticas en el uso del casco y de la bicicleta, y también en el número de víctimas mortales. Aunque sus resultados confirman que la legislación de uso obligatorio del casco reduce las víctimas mortales e incrementa el uso de esta medida de protección, estas medidas también parecen causar una reducción del uso de la bicicleta entre los jóvenes. Los autores mencionan como posibles mecanismos causantes de este resultado no intencionado el coste de los cascos, la existencia de alternativas más baratas que la bicicleta, y el estigma asociado a llevar un casco. Finalmente, y como consecuencia de todos esos efectos, sugieren que convendría realizar un análisis exhaustivo de la «relación coste-beneficio» antes de implantar este tipo de políticas.

2. Interacción entre distintas políticas y entre distintos niveles de Gobierno

La interacción entre políticas en distintos niveles de Gobierno también puede derivar en respuestas conductuales no intencionadas. Un ejemplo es el trabajo de Adams y Cotti (2008) sobre la legislación antitabaco a nivel estatal y local en Estados Unidos. Su trabajo usa como fuente de identificación causal la variación geográfica entre las prohibiciones de fumar locales y estatales. Contrariamente a lo esperado por los legisladores, observan un aumento en los accidentes mortales relacionados con la conduc-

ción bajo los efectos del alcohol. Este resultado, sorprendente a primera vista, pero robusto a una variedad de especificaciones, parece venir explicado por el hecho de que una parte de la población fumadora reaccionó ante la legislación antitabaco en su condado conduciendo distancias más largas hasta condados limítrofes en los que sí estaba permitido fumar en bares. Sus conclusiones muestran la utilidad, al aprobar una política pública, de tener en cuenta otras políticas relacionadas como lo es, en este caso, el posible impacto en la conducción bajo los efectos del alcohol (por ejemplo, impuestos del alcohol y legislación sobre la tasa de alcohol en sangre).

En general, la conclusión que puede obtenerse de estos trabajos es que el éxito de una política pública depende de los contextos normativos en el que se aplica y en concreto de que exista coordinación entre distintos niveles de Gobierno y distintas políticas públicas. Por ejemplo, al implantar una política local, los responsables políticos han de considerar las políticas estatales existentes y la posible interacción entre ambas normativas. Asimismo, estos trabajos también ponen de manifiesto la importancia de tener en cuenta las políticas existentes en territorios geográficamente limítrofes, las cuales podrían proporcionar una vía para que los ciudadanos eludan el cumplimiento de la nueva norma, haciéndola ineficaz y dando lugar, en ocasiones, a resultados contraproducentes.

3. Efectos de sustitución y arrastre

Al igual que ocurre con las conductas descritas en el apartado anterior, al implementar

políticas sanitarias eficaces han de tenerse en cuenta una multitud de efectos de sustitución y arrastre. Los Gobiernos deben ser conscientes de estos efectos, especialmente al implantar políticas destinadas a restringir el consumo de sustancias perjudiciales para la salud. Un ejemplo muy conocido es el estudio de DiNardo y Lemieux (2001), que analiza el impacto del incremento de la edad mínima para consumir alcohol sobre el consumo de bebidas alcohólicas y marihuana en alumnos estadounidenses del último curso de instituto. Los autores comparan distintos momentos del tiempo en los que gradualmente se aumentó la edad mínima de consumo de alcohol desde los 18 a los 21 años en diversos estados del país entre 1980 y 1989. Los datos utilizados proceden de una amplia muestra de alumnos extraídos de la encuesta *Monitoring the Future (MTF*, «Observando el futuro»). Los resultados de este estudio muestran de manera robusta que, si bien el aumento de la edad mínima redujo el consumo de alcohol, también incrementó el consumo de marihuana entre los estudiantes. En concreto, el destacado efecto positivo en términos estadísticos sobre el consumo de marihuana es sistemáticamente mayor en valores absolutos que el efecto negativo sobre el consumo de alcohol en todos los modelos (3,7 por 100 frente al 2,6 aproximadamente, en función de la especificación y las variables observables incluidas). No obstante, los autores señalan que dado que el porcentaje de estudiantes que consumen alcohol (67 por 100) es mucho mayor que en el caso de la marihuana (25 por 100), en realidad el efecto agregado es mayor para las bebidas alcohólicas que en la marihuana.

Este estudio es un ejemplo de cómo la consecuencia no intencionada de esta medida (aumento del consumo de marihuana) podría deberse a simples efectos de sustitución.

Las conclusiones del estudio anterior han sido matizadas posteriormente por otros trabajos que muestran que además del posible efecto sustitución también puede darse un efecto arrastre entre el alcohol y la marihuana. A este respecto, Crost y Guerrero (2012) constatan que si bien el consumo de alcohol en EE. UU. aumenta, como es de esperar, a los 21 años, el de marihuana también lo hace en torno a la misma edad. Es decir, cuando la normativa sobre bebidas alcohólicas se vuelve más permisiva (a los 21 años) y ambos productos están potencialmente disponibles, los resultados indican que el alcohol y la marihuana podrían ser productos complementarios. Por otro lado, Anderson, Hansen y Rees (2013) analizan los efectos que tiene la legislación sobre marihuana medicinal (LMM), que elimina las sanciones estatales por el consumo, posesión y cultivo de marihuana, en las muertes en accidentes de tráfico en Estados Unidos. Observan que, aunque la LMM reduce el número de víctimas mortales en carretera (aproximadamente entre un 8 y un 11 por 100 tras un año de su entrada en vigor), los efectos son mayores en las muertes relacionadas con el consumo de alcohol (disminución del 13,2 por 100), especialmente de los casos en los que al menos un conductor registra una tasa de alcoholemia superior al 0,10 (disminución del 15,5 por 100). Esto también respaldaría, en cierto modo, los potenciales efectos de interacción entre el alcohol y la marihuana.

A su vez, Carpenter y Dobkin (2011) examinan una variedad de consecuencias de la legislación que impone una edad mínima para consumir alcohol. En concreto analizan los efectos sobre el consumo de alcohol, la mortalidad, las lesiones no mortales y la delincuencia utilizando una amplia variedad de bases de datos estadounidenses. Concluyen que limitarse únicamente a reducir la edad mínima legal para consumir alcohol probablemente daría lugar a un aumento de los problemas relacionados con el alcohol. También indican que no existen evidencias suficientes para evaluar, ni por separado ni en conjunto, los efectos de políticas complementarias a las de establecimiento de una edad mínima, como pueden ser aquellas que educan sobre las consecuencias del consumo de alcohol, o las que afectan a la concesión de licencias para su venta. Por tanto, harían falta más datos y pruebas para determinar con certeza el efecto neto de las iniciativas relativas al alcohol.

Por último, las políticas relacionadas con el alcohol también pueden tener efectos *positivos* no intencionados. Un estudio reciente de Wada *et al.* (2017) emplea un modelo estructural que simula los efectos de un aumento del 5 por 100 en el IVA de las bebidas alcohólicas utilizando datos de Estados Unidos. Los autores constatan que, aunque subir los impuestos a las bebidas alcohólicas provocaría un desplazamiento de puestos de trabajo entre sectores de la economía (desde los sectores relacionados con el consumo de alcohol a los sectores no relacionados), el impacto en el número total de puestos laborales sería positivo, aunque escaso. Esto implicaría

que, aparte de los efectos intencionados sobre el consumo de alcohol según demuestran varios estudios (por ejemplo, Xu y Chaloupka, 2011), una subida de impuestos de este tipo también podría tener un efecto positivo general en el empleo.

En resumen, los resultados que hemos descrito acerca del impacto de las políticas sanitarias generan dudas sobre la eficacia global de algunas medidas de salud pública. Además, subrayan las ventajas que supone combinar *diversas fuentes de datos* que permitan analizar un *amplio rango* de resultados y que permitan el análisis de subgrupos poblacionales heterogéneos. Los Gobiernos que deseen reducir el consumo de productos perjudiciales para la salud, como el alcohol, mediante políticas que establezcan edades mínimas, prohibiciones e impuestos, no han de considerar los efectos en el consumo del producto en cuestión de manera aislada, sino que también deben tener en cuenta los posibles *efectos de sustitución* entre productos similares y los *efectos de arrastre* sobre otros sectores de la economía. Esta advertencia no solo se aplica a productos claramente perjudiciales para la sociedad como son el alcohol, la marihuana y el tabaco, sino también a cualquier producto para el que la legislación intente disuadir de su consumo. Por ejemplo, varios estudios acerca de los efectos indirectos de gravar con mayores impuestos los alimentos poco saludables y las bebidas azucaradas también recomiendan analizar detalladamente si las sustituciones podrían socavar los beneficios para la salud esperados de dichas medidas (por ejemplo, Cornelsen *et al.*, 2015). También se han observado con-

secuencias no intencionadas en materia de desigualdad, dado que la subida de impuestos de los alimentos poco saludables podría tener un carácter regresivo al favorecer a los consumidores con rentas más altas, que suelen consumir alimentos más saludables y tener mayor capacidad de respuesta ante las variaciones de precios (por ejemplo, Muller *et al.*, 2017).

III. EJEMPLOS DE POLÍTICAS EDUCATIVAS

En esta sección continuamos con la revisión de los efectos no intencionados de determinadas políticas públicas utilizando en este caso como ejemplo diversas medidas de política educativa que han dado lugar a resultados no previstos inicialmente.

1. Comportamientos no anticipados inicialmente

Nuestro primer ejemplo procede de la extensa literatura existente sobre lo que definiremos, en sentido amplio, como «programas de rendición de cuentas». Dos ejemplos de este tipo de políticas, provenientes de Estados Unidos, son el programa *No Child Left Behind* (NCLB, «Que ningún niño se quede atrás»), implantado durante la administración Bush, y el programa *Race to the Top* («Carrera a la cima»), implementado durante la administración Obama y que también incluye el uso de parámetros para la evaluación y remuneración de centros educativos y personal docente, como son medidas del valor añadido del profesor y la medición del progreso anual de cada centro. A nivel de centros educativos, el objetivo de estas políticas es mejorar el rendimiento educati-

vo, para lo que se utilizan datos sobre el rendimiento académico del alumnado en exámenes externos para premiar o castigar a los centros en función de su cumplimiento de las expectativas. El argumento es que, si existe una capacidad limitada de realizar un seguimiento eficaz de la calidad del centro y de su producción educativa, los docentes podrían comportarse de manera contraria a los intereses de la comunidad (esto es, padres o responsables políticos locales).

En principio, proporcionar información independiente y fiable acerca del buen funcionamiento de los centros, y asociar consecuencias al incumplimiento de las expectativas por parte del centro, puede incentivar a los docentes a ajustarse a las metas educativas establecidas (véase Figlio y Loeb, 2011). Sin embargo, poner en práctica la rendición de cuentas en los centros trae consigo una serie de decisiones prácticas; y es en estas decisiones donde surgen las consecuencias no intencionadas. Por ejemplo, para implementar una política de este tipo es necesario tomar una decisión normativa en cuanto a qué contenido incluir en los exámenes estandarizados, pues se plantea la disyuntiva de centrar la evaluación del alumno en un conjunto reducido de asignaturas o en uno más amplio. Por un lado, evaluar un conjunto amplio de asignaturas periódicamente tiene la ventaja de representar de manera más precisa una variedad de materias socialmente importantes, aunque al mismo tiempo puede ser difícil de cuantificar o puede tener un coste prohibitivo. Por otro lado, administrar estos exámenes resta tiempo adicional a la docencia lo que puede tener efectos en el rendimiento académ-

mico, por lo que, aunque existe cierta heterogeneidad en los programas de rendición de cuentas estadounidenses, la mayoría se centran más específicamente en la lectura y las matemáticas (Figlio y Loeb, 2011).

Desafortunadamente, está ampliamente documentado que cuando los exámenes son decisivos para la carrera del estudiante (*high-stakes testing*), algo que suele ser inherente a los programas de rendición de cuentas, estos provocan un fenómeno llamado «enseñar para el examen» (*teaching to the test*), en el que los centros educativos adaptan la formación para centrarse más activamente en las asignaturas evaluadas y menos en las que no se evalúan (por ejemplo, Ladd y Zelli, 2002; Stecher *et al.*, 2000; Linn, 2000). Poner el foco en las asignaturas evaluadas podría tener un efecto positivo si dichas asignaturas no recibieran la atención suficiente y representaran adecuadamente los objetivos de aprendizaje más importantes. Sin embargo, esta no parece ser la tónica dominante, puesto que la frase «enseñar para el examen» ha adquirido en sí misma una connotación peyorativa en las políticas educativas. Un enfoque limitado a unas pocas materias podría descuidar el tiempo dedicado al aprendizaje de conocimientos y habilidades valiosos, pero más difíciles de medir, como el razonamiento crítico o la creatividad. De la misma manera, incluso en las asignaturas evaluadas se ha observado que los docentes se limitan a impartir el material que se espera que aparezca en el examen, perdiendo así algunas de las competencias generales en dichas materias (por ejemplo, Koretz y Barron, 1998; Figlio y Loeb, 2011). También se ha documentado otro compor-

tamiento estratégico por parte de los docentes consistente en dedicar tiempo de enseñanza a mejorar la habilidad para realizar exámenes en lugar de dedicarlo al contenido curricular, así como centrar la formación únicamente en las áreas más fácilmente mejorables (Chakrabarti, 2013). También se han documentado comportamientos todavía más llamativos como los analizados en el trabajo de Jacob y Levitt (2003) en el que se constata que algunos docentes de Chicago cometieron fraude en los exámenes oficiales, llegando a modificar las respuestas de los alumnos para alterar los resultados académicos, lo que, lógicamente, repercutía también, en los resultados obtenidos por sus centros en los programas de rendición de cuentas.

Otro comportamiento estratégico que surge como respuesta directa a las presiones propias de los programas de rendición de cuentas de los centros educativos es la manipulación de a quién incluir en los exámenes. Resulta pertinente que en las políticas públicas se cuestione a quién incluir o excluir de los exámenes decisivos. Por ejemplo, ¿un centro debe rendir cuentas por los alumnos recién transferidos, por aquellos con discapacidades o por los que no sean hablantes nativos de la lengua de docencia? Han de establecerse normas para saber cómo contabilizar a dichos estudiantes con el fin de que los resultados de los exámenes sean fiables, aunque también se ha abusado de tales normas con fines estratégicos. Por ejemplo, se ha encontrado evidencia de que algunos centros educativos han reclasificado a los alumnos con bajo rendimiento académico como estudiantes con discapacidad para manipular la lista de examinados (por ejem-

plo, Figlio y Getzler, 2007; Cullen y Rebak, 2006), mientras que otros centros han modificado sus medidas disciplinarias particularmente durante los períodos de evaluación para que los alumnos con bajo rendimiento se quedaran en casa y, por tanto, no computaran en los promedios de los exámenes (Figlio, 2006).

Un objetivo adicional de estos programas de rendición de cuentas suele ser la mejora de los resultados académicos de los grupos poblacionales más desfavorecidos. No obstante, en muchas ocasiones, los modelos de evaluación empleados en los programas de rendición de cuentas no capturan de manera precisa las causas potencialmente importantes del bajo rendimiento académico de dichos grupos. Un sistema de evaluación que no permita tener en cuenta los problemas propios del perfil del estudiante ni la falta de recursos en centros más desfavorecidos atribuirá injustamente el incumplimiento de las expectativas al centro y a los docentes, que resultarán además injustamente castigados por el sistema, lo que provoca efectos adicionales contraproducentes (Ladd y Walsh, 2002). Si la falta de recursos escolares es el motivo del bajo rendimiento académico del alumno, un castigo que reduzca los recursos tan solo agravaría el problema. Figlio (2003) ha documentado algunas de estas consecuencias de financiación en distritos con muchas minorías y alumnos de pocos recursos en lo relativo a la normativa para el uso de los fondos incluidos en el programa *NCLB*.

Aunque sigue habiendo muy pocos datos acerca de los programas de mejora escolar en el marco de *Race to the Top* en

comparación con la amplia literatura existente sobre anteriores programas de rendición de cuentas, estudios recientes del estado de Carolina del Norte han constatado que la medida fue infructuosa y que, de hecho, es posible que haya mermado el rendimiento académico del alumnado, al haber restado tiempo de docencia por añadir carga administrativa, y haber derivado en rotaciones no anticipadas entre colegios de directores y profesorado (Heissel y Ladd, 2018). Una política que incentive cambios en la manera de enseñar en los centros educativos también modifica, evidentemente, las condiciones laborales del profesorado y, en consecuencia, puede alterar sus preferencias respecto a dónde trabajar. Nótese que esto no implica necesariamente que la rotación produzca efectos negativos, sino que podría tenerlos. De hecho, los efectos de esa rotación en la calidad docente y en el rendimiento del alumnado parecen ser heterogéneos (por ejemplo, Rockoff *et al.*, 2012; Clotfelter *et al.*, 2004).

En general, las conclusiones extraídas de los programas de rendición de cuentas que hemos resumido destacan la importancia de anticipar comportamientos estratégicos provocados por los incentivos incorporados en las políticas educativas. También muestran que un parámetro destinado a la medición, si se convierte en un parámetro que acarrea consecuencias académicas o laborales, puede alterar el comportamiento en formas deseadas y no deseadas. En general, el efecto neto de los programas de rendición de cuentas parece positivo, ya que dan lugar a un aumento en gran medida del rendimiento medio del alumnado en las asigna-

turas evaluadas tanto a corto plazo como a lo largo del tiempo (Figlio y Loeb, 2011). Sin embargo, pese a que los incentivos logren los objetivos marcados, se ha de ser consciente de que dichos incentivos también producen efectos negativos que limitan los beneficios de la medida.

2. Problemas asociados a la incorrecta generalización de resultados

Otro tipo de trabajos que arroja luz acerca de consecuencias no intencionadas de las políticas públicas son las investigaciones sobre el impacto académico del número de alumnos por aula. Al realizar una evaluación de esta medida, lo idóneo es que el investigador emplee algún tipo de variación exógena en el número de alumnos por aula, puesto que el tamaño de las clases es un factor endógeno en el que participan los propios centros y que tiene efectos en el alumnado que opta por un colegio u otro. Por ejemplo, podría darse el caso de que los centros con una menor concentración de alumnos por aula logren mejores resultados en los exámenes debido a que son muy selectivos o porque disponen de más recursos económicos, lo cual podría en ambos casos repercutir directamente en el rendimiento académico medio del alumnado. De la misma manera, si se asignan a propósito clases de menor tamaño a los programas dirigidos a determinados subgrupos de estudiantes, la relación entre el número de alumnos por aula y las calificaciones del alumnado en el entorno académico podría ser un simple reflejo de la selección del alumnado admitido en dichos programas (por ejemplo, cursos de nivel avanzado o programas de apoyo académico). Si los investi-

gadores o responsables políticos observan una correlación entre el número de alumnos por aula y el rendimiento académico del alumnado y lo consideran erróneamente un indicativo de que existe una relación causa-efecto entre ambas variables, esta incorrecta interpretación podría llevar a decisiones normativas equivocadas. Por ello, los trabajos más rigurosos sobre la cuestión suelen buscar variaciones exógenas en el número de alumnos por aula; esto es, variaciones en el tamaño de las clases que no guarden relación con factores que puedan repercutir en el rendimiento académico.

Un influyente estudio en esta línea de investigación evalúa el conocido proyecto STAR del estado de Tennessee, cuyas siglas en inglés significan «relación de rendimiento alumno/profesor» (student/teacher achievement ratio), que puso en marcha un experimento a gran escala en el que alumnos y profesores de educación infantil fueron asignados aleatoriamente a aulas de distintos tamaños y evaluados a lo largo del tiempo. Gracias a su diseño experimental y al gran tamaño de la muestra, este estudio se ha considerado un importante referente sobre los efectos de la reducción del número de alumnos por aula. Los datos sobre el impacto de esta política muestran que los resultados en los exámenes estandarizados fueron mejores en las clases de menor tamaño y que el beneficio fue mayor entre grupos desfavorecidos y con menos recursos económicos (Krueger, 1999). Asimismo, se ha observado que los efectos de este programa pueden ser duraderos. En este sentido, existe evidencia de que asistir a clases menos numerosas a edades tempranas incrementó

las probabilidades de que los alumnos pertenecientes a minorías se presentaran a pruebas de acceso a la universidad (Krueger y Whitmore, 2001).

También existen trabajos que estiman el efecto causal de las clases menos numerosas utilizando escenarios cuasiexperimentales. Estos trabajos llegan a conclusiones similares, ya que observan efectos positivos en las clases menos numerosas, particularmente para alumnos desfavorecidos (por ejemplo, Angrist y Lavy, 1999; Rivkin, Hanushek y Kain, 2005). No obstante, también existe evidencia sobre casos en los que la reducción del número de alumnos por aula en centros educativos (como es el caso del estado de Connecticut, estudiado en Hoxby [2000]) no parece tener un efecto causal.

Todos estos estudios tienen en común que aíslan el efecto causal del número de alumnos por aula en el contexto de estudio en cuestión. Es decir, analizan el impacto del tamaño de las clases *manteniendo todo lo demás constante*. Son investigaciones relevantes, puesto que pueden atribuir claramente sus efectos estimados exclusivamente al número de alumnos por aula. Sin embargo, en la aplicación real de tales políticas, esta estimación causal no es suficiente para saber qué repercusiones tendrá una potencial política en el alumnado, puesto que es muy improbable que el resto de factores educativos permanezcan invariables.

Un ejemplo ilustrativo de cómo la implementación a gran escala puede producir resultados contrarios a los documentados en estudios experimentales es la evaluación de Jepsen y Rivkin

(2009) del programa Class Size Reduction (CSR, «Reducción del número de alumnos por aula») en California. La implementación de un programa a gran escala de reducción del número de alumnos por clase implicará necesariamente la contratación de más profesorado y dicho efecto en el mercado laboral docente puede causar consecuencias no intencionadas. La iniciativa CSR de California, implementada en 1996, redujo el número de alumnos de 30 a 20 en las clases de dicho estado desde educación infantil hasta tercero de primaria. Jepsen y Rivkin (2009) observan que esta política aumentó ligeramente las calificaciones obtenidas en matemáticas y lectura, aunque los efectos de la reducción del número de alumnos por clase se vieron contrarrestados en gran medida por la disminución de la calidad del profesorado. El motivo de ello es que tal reducción a gran escala de alumnos por clase requiere una repentina contratación, también a gran escala, de personal docente, lo cual llevó a contratar profesores sin la debida titulación ni experiencia docente previa.

Detrás de estos resultados inesperados del programa CSR, se encuentra el hecho de que *no todos los demás factores se mantienen constantes en un contexto normativo a gran escala*. En este caso, disminuyó el número de alumnos por aula, pero también lo hicieron las cualificaciones del personal docente. Este es un importante efecto secundario de la política a tener en cuenta. En este sentido, Hanushek (2011) demuestra que el *efecto profesor* es enorme en el rendimiento académico y que la calidad docente es posiblemente uno de los aspectos más importantes para el aprendizaje del alumnado. El mencionado trabajo de

Hanushek puso de manifiesto que «reemplazar entre un 5 y un 8 por 100 del personal docente de peor calidad por profesores de calidad media podría posicionar a Estados Unidos cerca de los mejores puestos en los ránquines internacionales de matemáticas y ciencias, lo que supondría un valor actual de 100 billones de dólares» (Hanushek, 2011: p. 466). Comprender que en la práctica existe un equilibrio al menos a corto plazo entre la calidad docente y las clases menos numerosas pone de manifiesto que: i) apoyar la implementación de una política en los efectos causales que se han estimado asumiendo la condición *ceteris paribus* tan solo fundamenta parcialmente las decisiones normativas; y que ii) en los casos donde los efectos en el profesorado son incluso mayores que los efectos de la reducción del número de alumnos, se podría observar que una política bien intencionada acabe resultando contraproducente.

Cabe añadir que en el estudio de Jepsen y Rivkin (2009) acerca del programa CSR de California, los autores observaron una mayor proporción de docentes sin experiencia concentrados en los centros educativos con alumnado de estratos socioeconómicos más bajos. Esto se debe a que la contratación se vuelve aún más complicada en los centros más desfavorecidos (por ejemplo, Boyd *et al.*, 2013), por lo que cuando se convoca una gran oferta de plazas, existe tanto la necesidad de contratar a profesores nuevos y menos preparados para cubrir esa demanda, como también la necesidad de sustituir profesorado debido al traslado de algunos docentes de mayor cualificación desde entornos escolares más desfavorecidos

a centros más aventajados en los que haya plazas vacantes.

Otro estudio más reciente acerca de una política a gran escala de reducción del número de alumnos por clase en Florida también muestra resultados desalentadores para una inversión educativa tan elevada (Chingos, 2012). Chingos (2011) revisa la bibliografía acerca de la reducción del número de alumnos por aula y señala que, aunque puedan tener algunos beneficios muy concretos, por lo general, «estas políticas a gran escala también son extremadamente costosas y suponen una pérdida de oportunidades para realizar inversiones educativas más inteligentes» (Chingos, 2011: p.2).

En definitiva, los ejemplos de las políticas de California y Florida ponen de relieve la importancia de comprender las limitaciones de los resultados de un estudio científico a la hora de emplearlos para promover proyectos normativos.

3. Problemas asociados a la doble función de los centros escolares como centros educativos y como servicios de cuidado infantil

Un último ejemplo de la necesidad de prever correctamente los efectos secundarios de determinadas políticas proviene de los trabajos sobre la relación entre el mercado laboral y la doble función que desempeña a nivel familiar la educación de los niños pequeños. Además de proporcionar educación a los niños, el colegio también representa la principal fuente de cuidado infantil durante los días laborables. En consecuencia, las políticas que alteran el horario escolar

también alteran el horario del cuidado de los hijos y la jornada laboral de los padres y madres que trabajan.

Por ejemplo, en un intento de estimar cómo afecta el coste del cuidado de los niños en el mercado laboral de las madres, varios investigadores han considerado los programas de educación temprana como un sustituto de los servicios de guardería privados. En sus estimaciones, estas investigaciones utilizan la variación en la introducción de programas gratuitos de educación infantil y preescolar. Esto permite a los investigadores comparar a mujeres similares, algunas de ellas con la posibilidad de elegir una opción educativa pública gratuita y otras abocadas a pagar guarderías privadas para niños de la misma edad. Estos trabajos muestran que existen efectos positivos en el mercado laboral cuando el cuidado infantil se proporciona a través del sistema educativo, ya que ofrece una opción asequible para aquellas mujeres que se reincorporan al trabajo (por ejemplo, Gelbach 2002; Cascio 2009).

Otro ejemplo relevante son los cambios en el calendario escolar. Los calendarios escolares continuos modifican las fechas de asistencia a clase de los alumnos. Los colegios que adoptan estos calendarios suelen tener el mismo número de días lectivos que los colegios con calendarios normales, pero los redistribuyen a lo largo del año natural (Graves, McMullen y Rouse, 2013). Existen diferentes modelos sobre cómo implementar un calendario escolar continuo, pudiendo variar en la duración de los períodos lectivos y vacacionales alternos y en si todo el alumnado tiene el mismo horario de asistencia (*sin-*

gle-track; turno simple) o si se forman grupos de alumnos que vayan rotando según empiezan y terminan los períodos vacacionales (*multi-track*; turnos múltiples). Este tipo de calendarios suelen estar motivados por los supuestos beneficios académicos que conllevan y porque permiten combatir los problemas de saturación de la capacidad en un centro (véase Graves, McMullen y Rouse, 2013).

Desafortunadamente, los datos sobre el impacto académico no han resultado tan prometedores como se esperaba. En Carolina del Norte, McMullen y Rouse (2012a; 2012b) no observan que el cambio de calendario haya tenido un impacto académico relevante; tan solo constatan algunos efectos positivos en los centros con mayores problemas de aglomeración. Por otro lado, los datos de California indican que estos calendarios pueden afectar negativamente al aprendizaje de los alumnos, lo cual se ve aún más acrecentado en los grupos de alumnos desfavorecidos (Graves, 2010; 2011). Se ha observado que estos estudiantes desfavorecidos son quienes experimentan la mayor pérdida de conocimientos en verano, un fenómeno en el que los alumnos olvidan parte de lo aprendido en el anterior curso académico durante las extensas vacaciones de verano y en consecuencia a veces necesitan «ponerse al día» durante el primer trimestre del siguiente curso (por ejemplo, Downey, Von Hippel y Broh, 2004). Los responsables políticos que promueven calendarios académicos continuos basándose en posibles beneficios académicos derivados de acortar las vacaciones no suelen tener en cuenta que el hecho de que las vacaciones estivales no des-

aparecen en un calendario continuo, sino que se redistribuyen en descansos más frecuentes a lo largo del curso. La cuestión pertinente de cara a la formación de los alumnos es, por tanto, qué resulta más beneficioso para el aprendizaje: un único descanso o parones y comienzos más frecuentes. Y, en este sentido, parece que la evidencia no es favorable a los calendarios continuos.

La política de alterar los calendarios escolares, del mismo modo que se observó en las investigaciones sobre la reducción del número de alumnos por aula, pone de manifiesto un ejemplo de interpretación errónea de lo que puede aportarnos la información ofrecida por estudios científicos para fundamentar futuras decisiones normativas (por ejemplo, calendarios escolares). Sin embargo, también representa un claro ejemplo de que ignorar la función que desempeñan los centros educativos en el cuidado infantil puede producir efectos adversos no intencionados. En este sentido, Graves (2013a) pone de manifiesto que los calendarios escolares continuos impactan de manera negativa en el trabajo materno y que las madres que tienen algún hijo en un centro público con calendario continuo y también hijos más pequeños son quienes experimentan mayores dificultades para adaptarse. Este dato no sorprende si se tiene en cuenta que es más fácil encontrar servicios de cuidado infantil diferentes a los centros escolares durante los meses de verano, gracias a los campamentos y a que los alumnos de institutos y universidades, que aportan mano de obra a los campamentos, no tienen clase en verano. En cambio, si alguien intenta encontrar opciones para cuidar a sus hijos durante unas

vacaciones cortas en otros períodos no dispondrá de esas opciones. Graves (2103b) observa que los efectos en el trabajo materno son mayores en las madres blancas que en las minorías étnicas, lo cual la autora atribuye a que las minorías dependen en mayor medida de la familia para el cuidado de los hijos. La autora señala que «la familia quizá se adapte mejor a una distribución diferente de las necesidades de cuidado infantil, lo que permitiría a las madres pertenecientes a grupos minoritarios contrarrestar algunos (aunque no todos) de los efectos negativos sobre el empleo» (p. 279).

La conclusión de que los horarios escolares repercuten en la capacidad de las madres para trabajar no es exclusiva de los estudios sobre calendarios académicos. Como hemos mencionado antes, la implantación de programas públicos de educación infantil influye positivamente en el empleo de las madres (Gelbach, 2002; Cascio, 2009). Otras políticas como la ampliación de los programas de educación infantil a tiempo completo en lugar de a media jornada (por ejemplo, Gibbs, 2014) y los cambios en los días de la semana en que los niños asisten a clase (Duchini y Van Effenterre, 2020) también han influido en el mercado laboral de las madres. No cabe duda de que modificar la distribución del calendario escolar, ya sea a lo largo del año, de la semana o del día, supone un cambio de disponibilidad en cuanto al cuidado de los niños. Por consiguiente, si la sociedad desea facilitar la inserción laboral de la mujer, la distribución (días y horas) del calendario escolar debe ajustarse lo mejor posible al calendario y horarios propios del mundo laboral.

IV. CONCLUSIONES APLICABLES A POLÍTICAS ACTUALES Y FUTURAS

En este trabajo hemos ilustrado, con ejemplos, varios mecanismos comunes que provocan que las políticas públicas tengan resultados no intencionados. Obviamente, hemos proporcionado una muestra de ejemplos que no abarca todos los posibles mecanismos que podrían producir consecuencias no intencionadas de las políticas públicas.

Indudablemente, una recomendación que se deduce de nuestro trabajo es que resulta necesario prestar atención a las consecuencias no intencionadas de las nuevas políticas que se planteen teniendo en cuenta las políticas actuales. Un ejemplo al que cabe aplicar esta recomendación son las medidas para hacer frente a la pandemia mundial producida por la COVID-19. Las medidas de confinamiento y distanciamiento físico son, sin duda, políticas de salud pública que tienen el objetivo expreso de frenar los contagios para salvar vidas. No obstante, numerosas publicaciones recientes ya han constatado que también provocan muchos efectos secundarios que también convendría tener en cuenta. Por ejemplo, se ha demostrado que las medidas para frenar la COVID-19 mediante el confinamiento y el distanciamiento físico han provocado diversos efectos negativos en la salud mental, entre ellos depresión, ansiedad y un nivel de satisfacción con la vida cambiante a lo largo de las diversas olas de la pandemia, lo cual se observa en mayor medida entre la población joven y las mujeres (por ejemplo, Banks y Xu, 2020; Banks, Fancourt y Xu, 2021). En gene-

ral, es posible que el impacto en la salud mental de las políticas relacionadas con la COVID-19 haya acentuado las desigualdades de salud mental que ya existían por motivos de género, edad o étnicos (por ejemplo, Proto y Quintana-Domeque, 2021).

También se han observado efectos secundarios de las medidas de confinamiento y distanciamiento físico en el empleo y la educación. Se han documentado desigualdades raciales y de género en el desempleo (por ejemplo, Stevenson, 2020), así como también efectos sobre el reparto entre hombres y mujeres del cuidado de los hijos y las tareas domésticas, poniéndose de manifiesto que las mujeres que viven en entornos rurales han experimentado un aumento de la brecha de género tanto en el empleo como en términos de trabajo no remunerado, agudizando el fenómeno de la «doble jornada» (por ejemplo, Sevilla y Smith, 2020; Farré *et al.*, 2021). Asimismo, como los centros educativos han permanecido cerrados en algunos países en varias fases de la pandemia, la educación ha pasado a realizarse en el hogar o a distancia; y los datos preliminares muestran que esto también repercute en el aprendizaje de los estudiantes (Chatterji y Li, 2021; Musaddiq *et al.*, 2021).

No cabe duda de que estos ejemplos no aportan una visión completa de los efectos secundarios provocados por las medidas de política pública para frenar el avance de la COVID-19 (por ejemplo, cambios en los hábitos de consumo, impacto medioambiental, etc.), aunque sí que ilustran una esfera de acción actualmente en desarrollo donde investigadores y responsables políticos han de reflexionar dete-

nidamente acerca de todas las repercusiones que pueden tener estas medidas. También es importante destacar que aún es pronto para extraer conclusiones definitivas de muchas de las investigaciones en curso. Gran parte de la bibliografía documenta cambios, aunque no es capaz de identificar un efecto causal. En los casos en los que puede establecerse un efecto causal, este ha de interpretarse con cautela, puesto que obviamente no solo determina el impacto de una política, sino más bien el efecto conjunto de la medida política y también de las consecuencias directas de simplemente vivir en una pandemia. Por ejemplo, aunque las medidas de confinamiento por la COVID-19 tuvieron efectos en el empleo, no se sabe con certeza cuánto se habría resentido el mercado laboral si los trabajadores hubieran decidido abandonar sus puestos por miedo a la pandemia en caso de no haberse decretado un confinamiento. En las investigaciones a menudo se comparan los resultados del período previo a la pandemia y al confinamiento con los resultados durante el período tanto de pandemia como de confinamiento.

En definitiva, una de las principales contribuciones del presente trabajo consiste en poner de relieve que los diversos mecanismos descritos en este artículo pueden ser tenidos en cuenta para mejorar el diseño de futuros proyectos normativos, tanto en relación con la pandemia como en otras áreas. No obstante, es igualmente importante comprender y reconocer que muchos de los efectos secundarios de políticas analizados en este artículo pasaron desapercibidos para los investigadores durante varios años tras la entrada en vigor de aquellas. Este hecho pone de manifiesto

la importancia de realizar evaluaciones periódicas de las políticas públicas, así como la necesidad de establecer alianzas entre investigadores y responsables políticos en las primeras etapas de su implementación con el fin de poder identificar y rectificar cualquier consecuencia desfavorable en una fase temprana del proceso.

NOTA

(*) Los autores agradecen el apoyo financiero de la Línea de Estímulo de Jóvenes Doctores a través del Convenio Plurianual entre la Comunidad de Madrid y la Universidad Autónoma de Madrid (referencia: SI1-PJI-2019-00326).

BIBLIOGRAFÍA

- ADAMS, S. y COTTI, C. (2008). Drunk driving after the passage of smoking bans in bars. *Journal of Public Economics*, 92(5-6), pp. 1288-1305.
- ADDA, J. y CORNAGLIA, F. (2010). The Effect of Bans and Taxes on Passive Smoking. *American Economic Journal: Applied Economics*, 2(1), pp. 1-32.
- ANDERSON, M. D., HANSEN, B. y REES, D. I. (2013). Medical marijuana laws, traffic fatalities, and alcohol consumption. *The Journal of Law and Economics*, 56(2), pp. 333-369.
- ANGER, S., KVASNICKA, M. y SIEDLER, T. (2011). One last puff? Public smoking bans and smoking behavior. *Journal of Health Economics*, 30(3), pp. 591-601.
- ANGRIST, J. D. y LAVY, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2), pp. 533-575.
- BANKS, J., FANCOURT, D. y XU, X. (2021). Mental health and the COVID-19 pandemic. *World Happiness Report 2021*, pp. 107-130.
- BANKS, J. y XU, X. (2020). The Mental Health Effects of the First Two Months of Lockdown during the COVID-19 Pandemic in the UK. *Fiscal Studies*, 41(3), pp. 685-708.
- BOYD, D., LANKFORD, H., LOEB, S. y WYCKOFF, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31(1), pp. 83-117.
- CARPENTER, C. y DOBKIN, C. (2011). The minimum legal drinking age and public health. *Journal of Economic Perspectives*, 25(2), pp. 133-156.
- CARPENTER, C., POSTOLEK, S. y WARMAN, C. (2011). Public-place smoking laws and exposure to environmental tobacco smoke (ETS). *American Economic Journal: Economic Policy*, 3(3), pp. 35-61.
- CARPENTER, C. S. y STEHR, M. (2011). Intended and unintended consequences of youth bicycle helmet laws. *The Journal of Law and Economics*, 54(2), pp. 305-324.
- CORNELSEN, L., GREEN, R., DANGOUR, A. y SMITH, R. (2015). Why fat taxes won't make us thin. *Journal of Public Health*, 37(1), pp. 18-23.
- CASCIO, E. U. (2009). Maternal labor supply and the introduction of kindergartens into American public schools. *Journal of Human Resources*, 44(1), pp. 140-170.
- CHAKRABARTI, R. (2013). Vouchers, public school response, and the role of incentives: Evidence from Florida. *Economic Inquiry*, 51(1), pp. 500-526.
- CHATTERJI, P. y LI, Y. (2021). Effects of COVID-19 on school enrollment. *Economics of Education Review*, 83, p. 102128.
- CHINGOS, M. M. (2011). *The False Promise of Class-Size Reduction*. Center for American Progress.
- CHINGOS, M. M. (2012). The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate. *Economics of Education Review*, 31(5), pp. 543-562.
- CLOTFELTER, C., LADD, H., VIGDOR, J. y DÍAZ, R. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management*, 23(2), pp. 251-271.
- CROST, B. y GUERRERO, S. (2012). The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age. *Journal of Health Economics*, 31(1), pp. 112-121.
- CULLEN, J. y REBACK, R. (2006). Tinkering towards accolades: School gaming under a performance accountability system. En T. GRONBERG y D. JANSEN (eds.), *Advances in Applied Microeconomics*, 14 (pp. 1-34). Emerald Group Publishing Limited, Improving School Accountability: Check-ups or Choice.
- DE CHAISEMARTIN, C., GEOFFARD, P. Y. y LE FAOU, A. L. (2011). Workplace smoking ban effects on unhappy smokers. *Health Economics*, 20(9), pp. 1043-1055.
- DINARDO, J. y LEMIEUX, T. (2001). Alcohol, marijuana, and American youth: the unintended consequences of government regulation. *Journal of Health Economics*, 20(6), pp. 991-1010.
- DOWNEY, D. B., VON HIPPEL, P. T. y BROH, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, 69(5), pp. 613-635.
- DUCHINI, E. y VAN EFFENTERRE, C. (2020). School Schedule and the Gender Pay Gap. *IZA Discussion Papers*, 13791.
- FARRÉ, L., FAWAZ, Y., GONZÁLEZ, L. y GRAVES, J. (2021). Gender inequality in paid and unpaid work during COVID-19 times. *Review of Income and Wealth*. Forthcoming.
- FIGLIO, D. (2003). Fiscal implications of school accountability initiatives. *Tax Policy and the Economy*, 17, pp. 1-36.
- FIGLIO, D. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(4-5), pp. 837-851.
- FIGLIO, D. y GETZLER, L. (2007). Accountability, ability and disability: Gaming the system? En T. GRONBERG, D. JANSEN (eds.), *Advances in Applied Microeconomics*, 14, pp. 35-49. Emerald Group Publishing Limited, Improving School Accountability: Check-ups or Choice.
- FIGLIO, D. y LOEB, S. (2011). School Accountability. En A. ERIC, HANUSHEK, STEPHEN MACHIN y LUDGER WOESSMANN (ed.), *Handbooks in Economics*, Vol. 3. The Netherlands: North-Holland, pp. 383-421.

GELBACH, J. B. (2002). Public schooling for young children and maternal labor supply. <i>American Economic Review</i>, 92(1), pp. 307-322. GIBBS, C. R. (2014). A matter of time? Impact of statewide full-day kindergarten expansions on later academic skills and maternal employment. <i>The Department of Labor 2013-2014 Scholars Program</i>, p. 2. GRAVES, J. (2010). The Academic Impact of Multi-Track Year-Round School Calendars: A Response to School Overcrowding. <i>Journal of Urban Economics</i>, 67, pp. 378-391. GRAVES, J. (2011). Effects of Year-Round Schooling on Disadvantaged Students and the Distribution of Standardized Test Performance. <i>Economics of Education Review</i>, 30, pp. 1281-1305. GRAVES, J. (2013a). School Calendars, Child Care Availability and Maternal Employment. <i>Journal of Urban Economics</i>, 78, pp. 57-70. GRAVES, J. (2013b). The Effects of School Calendar Type on Maternal Employment across Racial Groups: A Story of Child Care Availability. <i>American Economic Review</i>, 103, pp. 279-283. GRAVES, J., McMULLEN, S. y ROUSE, K. (2013). Multi-Track Year-Round Schooling as Cost Savings Reform: Not Just a Matter of Time. <i>Education Finance and Policy</i>, 8, pp. 300-315. HANUSHEK, E. A. (2011). The economic value of higher teacher quality. <i>Economics of Education Review</i>, 30(3), pp. 466-479. HEISSEL, J. y LADD, H. (2018). School turnaround in North Carolina: A regression discontinuity analysis. <i>Economics of Education Review</i>, 62, pp. 302-320. HOXBY, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. <i>The Quarterly Journal of Economics</i>, 115(4), pp.1239-1285. JACOB, B. A. y LEVITT, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. <i>Quarterly Journal of Economics</i>, 118(3), pp. 843-878.	JEPSEN, C. y RIVKIN, S. (2009). Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. <i>Journal of Human Resources</i>, 44(1), pp. 223-250. JONES, A. M., LAPORTE, A., RICE, N. y ZUCCHELLI, E. (2015). Do public smoking bans have an impact on active smoking? Evidence from the UK. <i>Health Economics</i>, 24(2), pp. 175-192. KORETZ, D. M. y BARRON, S. (1998). <i>The validity of gains on the Kentucky Instructional Results Information System (KIRIS)</i>. Santa Monica, CA: RAND Corporation, 1998. KRUEGER, A. B. (1999). Experimental estimates of education production functions. <i>The Quarterly Journal of Economics</i>, 114(2), pp. 497-532. KRUEGER, A. B. y WHITMORE, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. <i>The Economic Journal</i>, 111(468), pp. 1-28. LADD, H. y WALSH, R. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. <i>Economics of Education Review</i>, 21(1), pp. 1-17. LADD, H. y ZELLI, F. (2002). School-based accountability in North Carolina: The responses of school principals. <i>Educational Administration Quarterly</i>, 38(4), pp. 494-529. LINN, R. L. (2000). Assessments and accountability. <i>Educ. Researcher</i>, 29(2), pp. 4-16. MULLER, L., LACROIX, A., LUSK, J. L. y RUFFIEUX, B. (2017). Distributional impacts of fat taxes and thin subsidies. <i>The Economic Journal</i>, 127(604), pp. 2066-2092. McMULLEN, S. C. y ROUSE, K. E. (2012a). School Crowding, Year-Round Schooling and Mobile Classroom Use: Evidence from North Carolina. <i>Economics of Education Review</i>, 31, pp. 812-823. McMULLEN, S. C. y ROUSE, K. E. (2012b). The Impact of Year-Round Schooling	on Academic Achievement: Evidence from Mandatory School Calendar Conversions. <i>American Economic Journal: Economic Policy</i>, 4, pp. 230-252. MUSADDIQ, T., STANGE, K. M., BACHER-HICKS, A. y GOODMAN, J. (2021). The Pandemic's effect on demand for public schools, homeschooling, and private schools. <i>Working Paper</i>, 29262. National Bureau of Economic Research. PROTO, E. y QUINTANA-DOMEQUE, C. (2021). COVID-19 and mental health deterioration by ethnicity and gender in the UK. <i>PloS one</i>, 16(1), p.e0244419. RIVKIN, S. G., HANUSHEK, E. A. y KAIN, J. F. (2005). Teachers, schools, and academic achievement. <i>Econometrica</i>, 73(2), pp. 417-458. ROCKOFF, J. E., STAIGER, D. O., KANE, T. J. y TAYLOR, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. <i>American Economic Review</i>, 102(7), pp. 3184-3213. SEVILLA, A. y SMITH, S. (2020). Baby steps: the gender division of childcare during the COVID-19 pandemic. <i>Oxford Review of Economic Policy</i>, 36(Supplement_1), pp. S169-S186. STECHEER, B., BARRON, S., CHUN, T. y ROSS, K. (2000). <i>The Effects of the Washington State Education Reform on Schools and Classrooms</i>. Center for Research on Evaluation, Standards and Student Testing. STEVENSON, B. (2020). The initial impact of COVID-19 on labor market outcomes across groups and the potential for permanent scarring. <i>The Hamilton Project</i>. Washington, DC: Brookings Institution. XU, X. y CHALOUKPA, F. J. (2011). The effects of prices on alcohol use and its consequences. <i>Alcohol Research & Health</i>, 34(2), p. 236. WADA, R., CHALOUKPA, F. J., POWELL, L. M. y JERNIGAN, D. H. (2017). Employment impacts of alcohol taxes. <i>Preventive Medicine</i>, 105, pp. S50-S55.
---	--	--

Resumen

Este artículo discute algunas líneas metodológicas para la evaluación de la eficiencia y la productividad de los servicios públicos. Para ello, se repasa la utilidad de realizar *benchmarking* mediante métodos frontera y se presentan tres desarrollos metodológicos recientes; metafronteras y tecnologías de referencia, evaluaciones causales de los programas para la mejora de la eficiencia y la evaluación de las ganancias potenciales de realizar fusiones de unidades que prestan un mismo servicio. Se concluye que para iniciar esta tarea sería necesario crear agencias autonómicas independientes de evaluación para aquellos servicios en los que la comunidad autónoma tiene competencias.

Palabras clave: eficiencia, productividad, frontera de producción, servicios públicos.

Abstract

The aim of this paper is introducing some methodology research lines for evaluating the efficiency and productivity of public services. To do this, I review the advantages of performing benchmarking using frontier methods together with three recent methodological developments; meta-frontiers and the use of reference technologies, causal evaluations of public programs to improve efficiency and the potential gains from merging units inside an industry providing the same service. It is concluded that to start this evaluation task it would be necessary to create Independent Evaluation Agencies at Autonomous Community level for those services in which the Autonomous Community has transferred powers.

Keywords: efficiency, productivity, production frontier, public services.

JEL classification: D21, D24, H11, H75.

DESARROLLOS RECIENTES EN EFICIENCIA Y PRODUCTIVIDAD PARA SU APLICACIÓN EN EL SECTOR PÚBLICO

Daniel SANTÍN GONZÁLEZ

Universidad Complutense de Madrid

I. INTRODUCCIÓN

LA crisis generada por la pandemia de la COVID-19 vuelve a poner en el centro del debate económico y político a unas administraciones públicas (AA. PP.) que se enfrentan a mayores demandas de servicios públicos por parte de unos ciudadanos cada vez más exigentes y mejor informados. Asimismo, el escenario macroeconómico indica que España cerrará 2021 con una deuda y un déficit público en el entorno del 120 por 100 y por debajo del 8 por 100 del PIB, respectivamente. Todo ello obliga, tal como ordena el artículo 31.2 de la Constitución Española, a redoblar esfuerzos para gestionar los impuestos pagados por los contribuyentes respondiendo a los criterios de eficiencia y economía.

Desde este punto de vista legal, las leyes presupuestarias de la mayor parte de países desarrollados también recogen el mandato de eficiencia en la gestión de los recursos públicos. En el caso español, el artículo 26.1 de la Ley General Presupuestaria, correspondiente a los principios y reglas de programación presupuestaria, señala que: «La programación presupuestaria se regirá por los principios de estabilidad presupuestaria, sostenibilidad financiera, plurianualidad, transparencia, eficiencia en la asignación y utilización de los

recursos públicos, responsabilidad y lealtad institucional (...)».

Para cumplir con esta voluntad del legislador, que no es otra que la voluntad que emana de los ciudadanos, la evaluación de la actividad del sector público constituye un proceso complejo, pero absolutamente imprescindible, para cuantificar la eficiencia de las distintas actividades que realizan las AA. PP. y poder así optimizar los servicios que se prestan a los ciudadanos. Podemos definir la evaluación del sector público como el proceso sistemático de generación de información e indicadores relevantes para establecer, mediante el análisis económico, una relación entre recursos, actividades y resultados con el objetivo de determinar si los programas y/o servicios públicos ofrecidos a los ciudadanos resultaron eficaces, eficientes y socialmente rentables. Todo ello con el objetivo de que tanto los políticos como la sociedad civil puedan usar los resultados en su toma de decisiones.

Por evidente que parezca al hablar de evaluación es conveniente volver a subrayar que no existe una única aproximación estadística, econométrica o matemática para llevar a cabo la evaluación de todos los programas, políticas y servicios que desarrolla el sector público. Así, no será lo mismo evaluar si merece la pena extender la alta velocidad para llegar a una nueva

ciudad, evaluar un programa para mejorar la empleabilidad de los jóvenes que abandonaron la educación secundaria y reducir así su vulnerabilidad en el mercado de trabajo, o la evaluación de la eficiencia de un conjunto de centros de salud que operan en una determinada región.

En las dos últimas décadas ha surgido un creciente interés por los experimentos controlados aleatorizados y las técnicas de inferencia que emplean una situación contrafactual para el análisis y la evaluación de programas y políticas públicas. El objetivo de toda esta positiva línea de trabajo es mostrar si existe una relación causal entre la intervención del sector público, mediante una política o programa determinado, y el cambio en el indicador o indicadores asociados a la población beneficiaria del programa con relación a un grupo de control correctamente definido. Prueba de este interés y de la robustez de los resultados que están siendo obtenidos con estas técnicas son las recientes concesiones en 2019 (Barnejee, Duflo y Kremer) y en 2021 (Angrist, Card y Imbens) del Premio Nobel de Economía a profesores que desarrollan estas líneas de trabajo.

Siendo del todo indiscutible la importancia de esta dimensión de análisis causal en la evaluación de políticas públicas conviene también volver a recordar la importancia de llevar a cabo la evaluación sistemática de la eficiencia y la productividad de los servicios públicos mediante el uso de fronteras productivas. La oferta de servicios públicos que se realiza mediante unidades independientes de producción pública, que consumen trabajo y capital para entregar servicios de

manera satisfactoria a una población determinada, es ingente.

Redondeando, y únicamente a título ilustrativo, en España podemos contar unos 4.500 centros donde se imparte Educación Infantil, 9.800 centros de Educación Primaria, unos 4.000 centros de Educación Secundaria Obligatoria y Bachillerato, unos 460 hospitales, 3.000 centros de salud, 10.000 consultorios médicos locales, unos 1.000 juzgados de primera instancia e instrucción, 230 comisarías de policía, 130 servicios de bomberos, 68 cárceles, 47 universidades públicas presenciales, 46 puertos, 45 aeropuertos y un largo etcétera.

El gasto público asociado a la prestación de estos servicios es muy elevado y, además, a diferencia de «programas para la mejora de» que empiezan y en algún momento acaban, se repiten de forma sostenida año tras año. En 2018, el gasto presupuestado por las administraciones públicas para hacer frente a los servicios sanitarios (educativos) fue de unos 77.000 (45.000) millones de euros (Ministerio de Sanidad, 2021; MEFP, 2021). Esto supone que, solo en sanidad y educación, por cada punto porcentual que pudiéramos mejorar la eficiencia en la gestión de estos servicios liberaríamos 1.220 millones de euros bien para usos alternativos o bien en reducción de impuestos.

En el contexto de la evaluación del sector público, el interés por mejorar su eficiencia ha constituido uno de los principales focos de atención e investigación por parte de la economía pública, así como por investigadores de distintos ámbitos. Por motivos de extensión, se considera que no tiene sentido

volver a repasar ni los conceptos económicos de eficiencia y productividad ni todas y cada una de las metodologías de fronteras productivas (1) estándares disponibles para medir la eficiencia, tanto no paramétricas, incluyendo el análisis envolvente de datos (*data envelopment analysis-DEA*) como paramétricas, siendo el análisis con fronteras estocásticas (*stochastic frontier analysis-SFA*) la técnica más aplicada. Para este fin se propone al lector los trabajos previos de Álvarez (2001), Coelli et al. (2005), Fried et al. (2008) y Kumbhakar, Wang y Horncastle (2015) (2).

Tampoco tendría sentido utilizar unas pocas páginas para, desde una perspectiva sectorial, hacer un repaso pormenorizado de las distintas actividades que produce el sector público y que son susceptibles de ser analizadas con especial atención al caso español. Para ello, se refiere al lector al excelente número 95 de *Papeles de Economía Española* (Funcas, 2003) que llevó por título «Sector Público y Eficiencia» y al número 105 de la *Revista Economistas* (Colegio de Economistas de Madrid, 2005) titulado «La eficiencia de los servicios públicos: viejos problemas, nuevos enfoques».

El objetivo de este artículo es discutir algunas de las líneas metodológicas que considero que podrían ser utilizadas para la evaluación de la eficiencia y la productividad en el ámbito de la prestación de los servicios públicos. Para ello, en la segunda sección se repasa la utilidad de medir la eficiencia y la productividad en el sector público y cuál es la situación actual en España respecto a este tipo de evaluación. En la sección tercera se presentan tres desarrollos me-

todológicos e ideas que podrían ser utilizadas para llevar a cabo el *benchmarking* de unidades productivas. Esto incluye la cuantificación de las diferencias en productividad mediante metafronteras y tecnologías de referencia, las evaluaciones causales de los programas para la mejora de la eficiencia y la evaluación de las ganancias potenciales de realizar fusiones y escisiones de unidades de producción pública. Finalmente, en la última sección se ofrecen conclusiones y líneas futuras de trabajo.

II. LA MEDICIÓN DE LA EFICIENCIA Y LA PRODUCTIVIDAD DE LOS SERVICIOS PÚBLICOS EN ESPAÑA

Cualquier manual de microeconomía define una tecnología como la capacidad de obtener distintos *outputs* a partir del uso de múltiples *inputs*. Su caracterización básica se realiza mediante el concepto de función de producción. Así, una función de producción $y = f(x)$ indica la *máxima cantidad* y de *output* que es factible producir a partir de una determinada cantidad de *inputs* $(x_1, \dots, x_N) > 0$ (Mas-Colell, Whinston y Green, 1995, p.129). Se subraya «máxima cantidad» porque rara vez los manuales de microeconomía definen, asociado al concepto de función de producción, el concepto de eficiencia.

En efecto, en mercados que operan en competencia perfecta y con productores privados precio-aceptantes guiados por la maximización del beneficio, no caben resultados ineficientes ya que estos llevarían directamente a la quiebra a estas empresas y, por tanto, se asume que la pro-

ducción se lleva a cabo con plena eficiencia. En principio, nada impediría que la producción de bienes y/o servicios públicos alcance idénticos resultados. Sin embargo, por distintas razones asociadas a la gestión de las unidades productivas las unidades de producción públicas encuentran dificultades para alcanzar los máximos teóricos permitidos por la tecnología (Farrell, 1957; Leibenstein, 1966; Levin, 1974; Niskanen, 1975; Levin, Jamison y Radner, 1976). La multiplicidad de objetivos heterogéneos y a veces no definidos, la falta de restricciones presupuestarias claras y la estructura de incentivos pueden apartar a la empresa pública de alcanzar la máxima eficiencia (Tirole, 1994).

Ello implica que es factible que las unidades pueden llevar a cabo la producción de los servicios públicos de manera ineficiente, esto es, utilizando más recursos de los que se requieren para lograr una cantidad de producción o , alternatively, dependiendo de la orientación elegida, produciendo por debajo del máximo tecnológicamente factible dados unos recursos. A partir de este hecho, y dado que rara vez se observan quiebras en los sectores enumerados en la sección anterior, surge la necesidad de conocer la magnitud de la eficiencia con la que se lleva a cabo la producción de un servicio público comparando a las distintas unidades productivas que lo desarrollan.

La utilidad de esta evaluación vista la magnitud presupuestaria que supone la prestación de estos servicios, es considerable. En este sentido, el australiano Steering Committee for the Review of Government Service Provision (SCRGP, 2013) señala que tra-

dicionalmente todos los esfuerzos de las administraciones públicas para mejorar los servicios se han centrado en incrementar los recursos presupuestarios, como si fuera evidente que más euros invertidos automáticamente fueran capaces de producir más y mejores servicios. Sin embargo, el objetivo de las AA. PP. debería cambiar y buscar las mejores maneras de utilizar los recursos públicos existentes (muchos o pocos, su magnitud, dependerá de las preferencias democráticas que desean los ciudadanos acerca del peso del sector público que quieren para la economía donde residen) desplazando el foco de atención desde la cuantía de los recursos, «más» no supone tautológicamente «mejor», a un uso de los mismos eficiente.

Las principales ventajas de llevar a cabo la evaluación del desempeño de los servicios públicos mediante competencia por comparación son numerosas, pero pueden destacarse las siguientes. En primer lugar, ayuda a definir el cumplimiento de objetivos, tanto cuantitativos como cualitativos, así como a clarificar las responsabilidades de las organizaciones. En segundo lugar, anima a promover el análisis de las relaciones entre agencias y programas para una mejor coordinación de las políticas dentro de las AA. PP. También mejora la transparencia, ofreciendo tanto a los distintos Gobiernos, central, autonómico y local, como a los ciudadanos indicadores de desempeño de sus políticas y servicios públicos a lo largo del tiempo. En tercer lugar, la competencia por comparación anima a la continua mejora en el desempeño ya que, conocidos los indicadores de desempeño, ninguna unidad productiva querrá

aparecer la última en el *ranking*. De esta manera el *benchmarking* permite identificar a las mejores prácticas, aprender de su gestión y promover la introducción de mejoras e innovación en la entrega de los servicios. Y no menos importante, también permite identificar a las peores prácticas, con el objetivo de poner remedio a los problemas que impiden que ofrezcan el servicio de manera eficiente, mediante el establecimiento de objetivos de mejora.

Hasta la fecha, la institucionalización de la evaluación sistemática de los servicios públicos en España ha tenido escaso éxito. El primer intento serio de llevar a cabo una evaluación de la eficiencia del gasto público tuvo su origen en la creación, a mediados de 2006, de la Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (AEVAL). La misión de AEVAL era «la promoción y realización de evaluaciones y análisis de impacto de las políticas y programas públicos, así como el impulso de la gestión de la calidad de los servicios, favoreciendo el uso racional de los recursos y la rendición de cuentas a la ciudadanía». Sin embargo, las evaluaciones planteadas por AEVAL no incluían los servicios prestados por las comunidades autónomas (CC. AA.), solo los de la Administración General del Estado, y no respondieron claramente a los objetivos definidos, por lo que la Agencia fue definitivamente clausurada en 2017.

En 2016, como consecuencia de la crisis financiera global, España adquirió el compromiso ante la Comisión Europea de revisar la eficiencia de las administraciones públicas. Esta tarea de análisis y revisión del gasto público conocida como *spending review*

es encargada por el Gobierno a la Autoridad Independiente de Responsabilidad Fiscal (AIReF) que había sido creada en 2014. El primer Plan de Actuación del *spending review* desarrollado por AIReF marcó un punto de inflexión al incluir, por primera vez de manera explícita en un texto de una institución pública, la posibilidad de realizar *benchmarking* mediante el uso de fronteras productivas, lo que AIReF (2017, p. 6) denomina «frontera eficiente». Concretamente, el Plan de Acción de AIReF afirma: «Todo ello con objeto de medir tanto el “coste medio” de la intervención pública como “gasto medio” por beneficiario. Estas evaluaciones también pueden incluir el cálculo de fronteras eficientes de producción, que determinan un *benchmarking*». Si bien estos análisis han empezado a aparecer en algunas de sus evaluaciones, como en AIReF (2018) para el análisis de la Sociedad Estatal Correos y Telégrafos o en AIReF (2020) para el estudio del sistema universitario público andaluz, se trata todavía de análisis muy incipientes que distan de ser sistemáticos (3). En la siguiente sección se apuntan algunas ideas para institucionalizar el *benchmarking* en las AA. PP.

III. LA EVALUACIÓN DE LA EFICIENCIA Y LA PRODUCTIVIDAD EN EL SECTOR PÚBLICO

Una vez realizadas las primeras aproximaciones con algunas evaluaciones puntuales, el siguiente paso es institucionalizar la evaluación del desempeño de la producción de los servicios públicos. Es decir, la evaluación de la eficacia, la eficiencia y la productividad de manera continuada en el tiempo, y que en

cada ejercicio explora la mejora de los mecanismos de búsqueda y análisis de los indicadores asociados a esta evaluación. Para ello, el primer requisito siempre debe ser la definición de los *inputs* y *outputs* que serán utilizados y que idealmente deberían incluir indicadores de cantidad, calidad (4), *outputs* no deseables (5) así como variables que tengan en cuenta el entorno en el que operan (6) las distintas unidades. En la mayoría de los casos la literatura es amplia y las posibilidades son múltiples, por lo que se recomienda alcanzar el mayor consenso posible entre evaluador, unidades evaluadas y usuarios finales de la evaluación teniendo en cuenta el contexto y los objetivos perseguidos. Ello también incluye la realización de distintos análisis para comprobar la robustez en los resultados.

A continuación, deben definirse las fuentes de datos de donde se obtendrá la información necesaria para ejecutar los análisis. En numerosas ocasiones las evaluaciones de eficiencia y productividad no se llevan a cabo por parte de las AA. PP. o por parte de otros agentes argumentando que no existe información acerca de todas las variables con las que sería necesario contar. En este sentido, es importante señalar que siempre existirá información e indicadores no disponibles y también siempre será posible perfeccionar las medidas ya existentes. Este hecho genera un círculo vicioso ya que los indicadores no se generan porque no existe evaluación y la evaluación nunca empieza porque se considera que no existen los indicadores adecuados. El resumen es que para romper este círculo es imprescindible la voluntad política de comenzar a evaluar.

Como ya se ha señalado, la evaluación es un proceso de generación de información relevante continuo en el tiempo y, por tanto, posiblemente la información para el análisis nunca será totalmente perfecta. Comenzar a evaluar supone diseñar sistemas de información para almacenar y gestionar los datos necesarios en la evaluación. Inicialmente estos serán imperfectos, pero con el paso del tiempo las medidas serán cada vez más precisas y adecuadas para conseguir mejores evaluaciones. Un ejemplo de la puesta en servicio de un sistema de información sería euroCRIS, (Current Research Information Systems for Europe), que recoge datos sobre la investigación llevada a cabo por distintas instituciones y universidades a partir de las actividades de sus profesores e investigadores.

En este ejemplo, una vez obtenida esta información, la misma puede ser utilizada para analizar la eficiencia y la productividad del sistema de investigación español en comparación con otros países europeos. Además, el sistema permitiría también comparar los centros de investigación en España. Por supuesto, la puesta en marcha de un sistema de información llevará tiempo, pero esta no parece una dificultad insalvable en la sociedad interconectada del siglo XXI.

Obtenida la información cabe plantearse cuáles son los objetivos de análisis que se pueden desarrollar. A continuación, señalo tres líneas de trabajo que, en mi opinión, serían especialmente deseables para mejorar la toma de decisiones: 1) las comparaciones y evolución de la productividad total de los factores; 2) la evaluación causal de programas

públicos para la mejora de la eficiencia y la productividad; y 3) el análisis de la fusión y escisión de unidades que prestan servicios públicos.

1. Metafronteras y diferencias en productividad a lo largo del tiempo

Como ya se ha señalado, la medición de la eficiencia con la que operan los servicios públicos es una parte fundamental para proporcionar los mejores servicios posibles a los ciudadanos haciendo un buen uso de los recursos. Además de los objetivos ya señalados, Pedraja-Chaparro, Salinas-Jiménez y Smith (1999) señalan como objetivo de esta evaluación «el análisis de la eficiencia global de la industria». En el caso del sector público español entendemos por industria al conjunto de unidades productivas que producen un servicio en un determinado sector que está bajo la jurisdicción de una misma administración (7). Algunos ejemplos podrían ser los institutos de Educación Secundaria (IES) de la Comunidad de Castilla y León, las oficinas de la empresa pública Correos o los hospitales andaluces. Estos centros forman «una industria» que depende presupuestariamente de una Administración interesada en ofrecer el mejor y más eficiente servicio a sus ciudadanos.

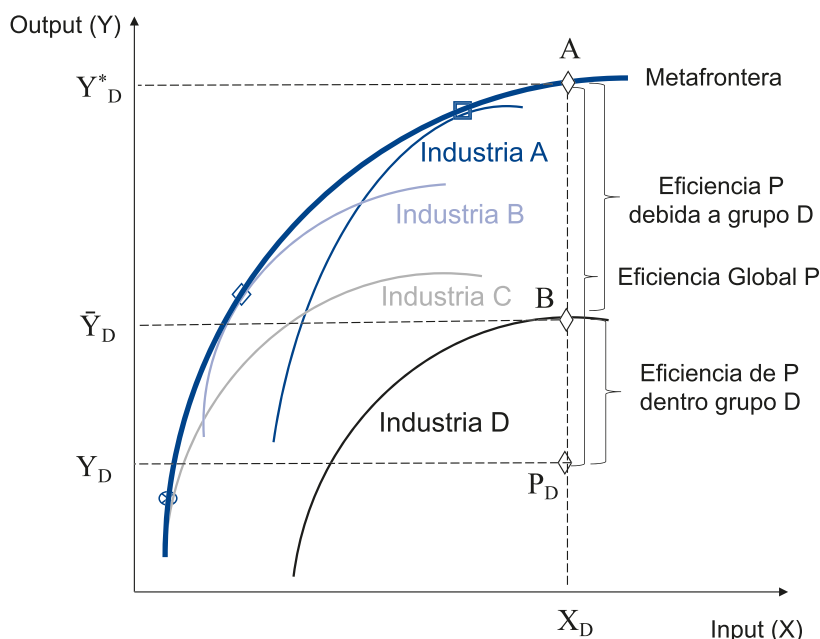
A partir de aquí el primer objetivo de estas administraciones es analizar la eficiencia de «su industria» y cómo evoluciona a lo largo del tiempo. A continuación, podemos comparar el desempeño de la industria con otras industrias semejantes. Por ejemplo, podemos analizar las diferencias en desempeño de los

IES con los de los centros de educación concertada de la misma CC. AA. que prestan con dinero público los mismos servicios. También podemos comparar el desempeño de grupos de centros que operan en municipios de distinto tamaño o incluso comparar el desempeño de médicos dentro de un mismo hospital.

Siendo estas primeras mediciones de la eficiencia y de los cambios en productividad a lo largo del tiempo importantes, el siguiente paso sería realizar el análisis de manera comparada no solo dentro de cada industria, sino entre industrias que pertenecen a distintas jurisdicciones. Que parezca que nuestra industria va bien será una medida vacía si nunca comparamos sus resultados con los que obtienen otras industrias de nuestro entorno (8). Podríamos entonces comparar industrias de distintas comunidades autónomas o incluso de distintos países a lo largo del tiempo.

Por tanto, se trata de recuperar la idea de «federalismo de laboratorio» como una importante ventaja del sistema autonómico español todavía no bien explotada. Para Oates (1999), un sistema descentralizado permite probar y experimentar distintas alternativas de prestar un servicio público. Esta experimentación tendría por objetivo obtener ganancias de eficiencia en términos de progreso tecnológico que a su vez animen a más experimentación. En definitiva, las CC. AA. podrían introducir distintas innovaciones en sus servicios para maximizar el nivel de los *outputs* que entregan a sus ciudadanos dado un nivel de gasto, al tiempo que reproducen los programas exitosos en otras CC. AA. y descartan aquellos cambios en

GRÁFICO 1
EL CONCEPTO DE METAFRONTERA



Fuente: Elaboración propia.

los programas que no suponen ganancias de eficiencia.

Para realizar este «federalismo de laboratorio» acerca de lo que funciona y lo que no funciona en términos de mejorar la eficiencia y la productividad de los servicios podemos recurrir al uso de metafronteras (para una revisión se recomienda las lecturas de Battese y Prasada-Rao, 2002; Pastor y Lovell, 2005 y Oh, 2010).

El concepto de metafrontera recoge la idea de trazar una única frontera a partir de las mejores prácticas de las unidades productivas pertenecientes a las distintas industrias y podemos ilustrarlo a partir del gráfico 1.

En el gráfico 1 se representan las fronteras de producción de cuatro industrias; A, B, C y D, capaces de producir un *output* a partir de un *input*, cada una de

ellas formada por un número de unidades de producción. La idea de metafrontera implica trazar una nueva frontera productiva asumiendo que la combinación lineal de las mejores unidades productivas pertenecientes a los distintos grupos o industrias es tecnológicamente factible. Así, la eficiencia de cualquier unidad productiva se podría descomponer en dos partes, una parte debida a su propia gestión en comparación con otras unidades productivas que compiten con ella en la misma industria y una parte debida a la diferencia tecnológica entre las industrias comparadas.

La unidad productiva no sería directamente responsable de esta segunda parte de la distancia a la frontera, sino que lo serían los técnicos y políticos encargados de la industria, que deberían analizar el porqué de

estas diferencias. Por un lado, las causas podrían ser exógenas, dependientes de variables geográficas, de dispersión de la población o en cualquier caso de factores fuera del alcance de los gestores de la industria. En este caso, existen distintas metodologías para incorporar directamente estas variables en el análisis. Finalmente, la parte no explicada por estos factores exógenos sería eficiencia en la gestión o cuellos de botella debidos a marcos regulatorios que impiden el pleno desarrollo de la industria para favorecer otros objetivos distintos de la eficiencia.

Por ejemplo, en el gráfico 1 P_D sería una unidad ineficiente ya que la frontera productiva construida para su industria D indica que con la cantidad de *input* que utiliza podría aumentar la cantidad actual que produce de *output* Y_D hasta el nivel Y_D-bar. Además, si comparamos a la industria D con la metafrontera, construida para los restos de unidades analizadas, observamos cómo existe un objetivo de producción potencial que se sitúa en Y_D*. Siendo la diferencia entre Y_D* y Y_D-bar debida a la industria. Las metafronteras han sido empleadas en España para el análisis de servicios locales de recogida de basura (Garrido-Rodríguez *et al.*, 2018), el transporte público (Campos-Alba, 2020) o para evaluar diferentes sistemas educativos (Cordero, Santín y Simancas, 2017; Giménez *et al.*, 2019), entre otros.

La metafrontera puede ser estimada mediante distintos procedimientos como una frontera estocástica, DEA u orden-m, pero un supuesto básico para el uso de esta metodología es asumir que la tecnología sintética construida en la metafrontera

es factible y reproducible en la industria que se desea mejorar. Esto en la práctica puede suponer problemas prácticos ya que las unidades productivas en la metafrontera no operan en los mismos grupos y por tanto la referencia es difícil de copiar ya que sus *DMUs* (*decision making units* o unidades de decisión) operan en distintas industrias bajo distintos marcos legales, contextos sociales o económicos, lo que puede llegar a hacer imposible alcanzar la referencia.

Para superar esta dificultad en la comparación de industrias, Camanho y Dyson (2006) proponen adaptar el índice de Malmquist (Caves, Christensen y Diewert, 1982) donde los dos períodos de tiempo que se utilizan en la definición de este índice se sustituyen por dos grupos de unidades productivas que son comparados. El objetivo de estos autores es superar esta dificultad de las metafronteras para hacer comparaciones directas de las diferencias en productividad entre

industrias. El gráfico 2 muestra un ejemplo para entender mejor esta metodología.

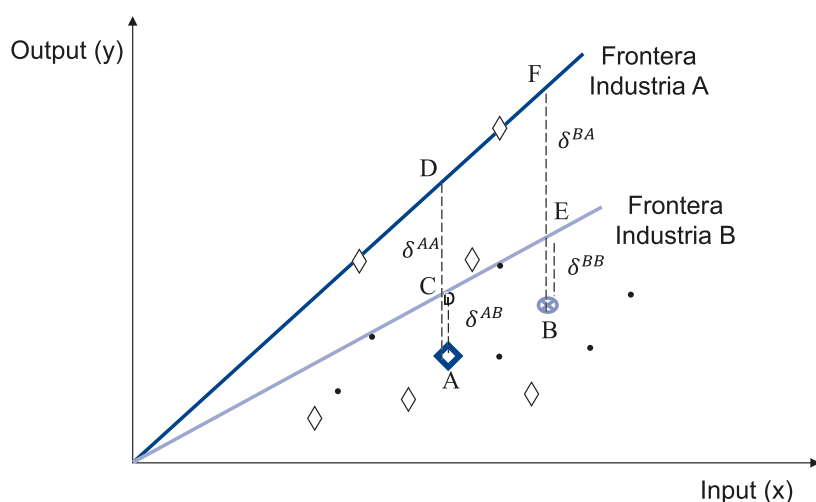
En el gráfico 2 observamos las fronteras productivas de dos industrias A y B cuyas unidades están representadas por rombos y círculos respectivamente. Por un lado, podemos ver cómo la frontera productiva de la industria A domina en términos tecnológicos a la de la industria B, ya que con la misma cantidad de *input* es capaz de obtener mayor *output*. Para cada unidad productiva podemos calcular este índice de Malmquist adaptado, como la diferencia en productividad entre dos industrias, A y B, en vez de entre dos períodos, t y $t + 1$ como haríamos normalmente con aquel índice.

En el gráfico 2 los puntos A y B representan la posición hipotética promedio de las dos industrias, determinadas a partir de las medias geométricas de las distancias de las unidades de cada industria respecto a las dos fronteras. Esto

nos permite descomponer las diferencias en productividad entre las dos industrias, como en el índice de Malmquist, en una parte debida a diferencias tecnológicas y otra a diferencias en eficiencia. Así, podemos comprobar cómo, aunque por causas tecnológicas (media geométrica de las distancias DC y EF) la productividad en A es mayor que en B, la productividad de la industria B es mayor que la A en términos de eficiencia (la distancia BE es menor que la distancia AD y, por tanto, está más cerca de la frontera). El producto de la diferencia en productividad debida a tecnología y a eficiencia nos dará una medida acerca de cuál es la diferencia global en productividad entre las industrias. Esta información es muy útil para aprender qué podemos hacer para mejorar la productividad del servicio que ofrecemos a los ciudadanos a partir de la comparación con otras industrias que también lo desarrollan.

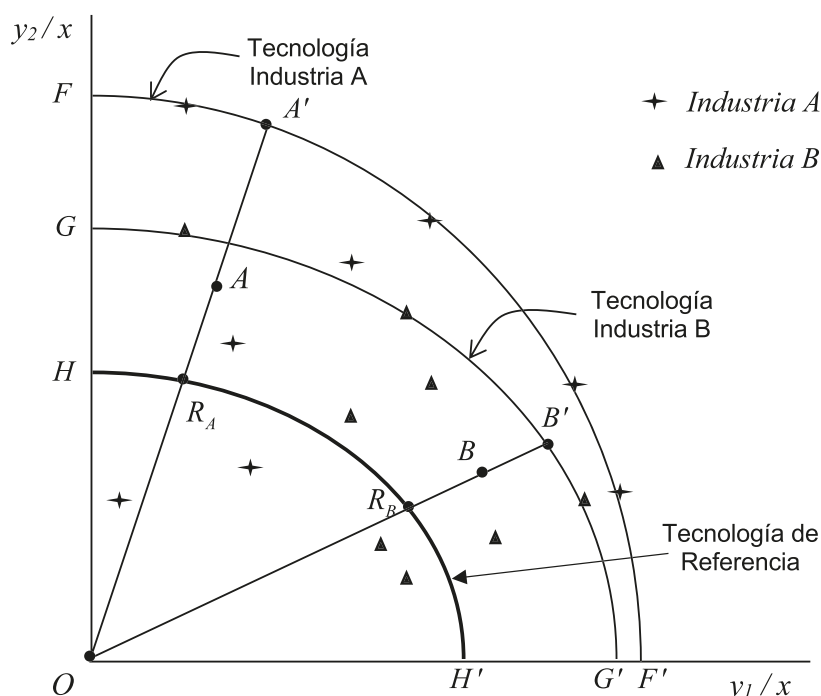
El índice de Camanho y Dyson es atractivo y ha sido utilizado para evaluar hospitales según la titularidad con la que actúan (Ferreira y Cunha-Marques, 2015) o la gestión de aeropuertos (Ferreira, Marques y Pedro, 2016). Sin embargo, el índice presenta dos inconvenientes para su uso sistemático en la evaluación del sector público. En primer lugar, no se pensó para hacer seguimientos de las diferencias a lo largo del tiempo. En segundo lugar, con más de dos grupos se rompe la condición deseable de circularidad de Frisch (1936) en números índices. Es decir, dados tres grupos A, B y C, la diferencia global en productividad, y en sus componentes entre los grupos A y C, debe ser el producto de las diferencias entre A y B por una parte y B y C por otra.

GRÁFICO 2
DIFERENCIAS EN PRODUCTIVIDAD ENTRE INDUSTRIAS
(CAMANHO Y DYSON, 2006)



Fuente: Elaboración propia.

GRÁFICO 3
DIFERENCIAS EN PRODUCTIVIDAD ENTRE INDUSTRIAS
(APARICIO Y SANTÍN, 2018)



Fuente: Elaboración propia.

Para solucionar este problema, Aparicio y Santín (2018) proponen un nuevo índice basado en Berg, Førsund y Jansen (1992) que consiste en introducir la tecnología de un «grupo de referencia» (R). Con ello, los problemas de circularidad entre dos o más grupos a lo largo del tiempo desaparecen. Otra ventaja de este índice es que no es necesario actualizar la metafrontera en cada nuevo período, lo que permite hacer un seguimiento sistemático y robusto de las diferencias de productividad entre industrias a lo largo del tiempo (9). El gráfico 3 muestra las ideas básicas y las ventajas de introducir una referencia para el cálculo de este índice.

En el gráfico 3 encontramos nuevamente dos industrias A y B que utilizan un único input

para producir dos *outputs*. Al igual que en los gráficos 1 y 2, el gráfico 3 muestra las fronteras productivas definidas por las tecnologías de cada industria («FF» y «GG» para las industrias A y B respectivamente). Pero, para realizar las comparaciones, en lugar de utilizar una metafrontera utilizamos la tecnología (HH') de un grupo de unidades productivas que las industrias que serán evaluadas consideran una referencia. La proyección de las medias geométricas de las unidades de cada sector, representadas en el gráfico 3 por los puntos A y B, respecto a su propia tecnología (AA» y BB') y respecto a la tecnología de referencia (AR_A y BR_B) permiten medir cómo las diferencias en productividad de las dos industrias evolucionan a lo largo del tiempo y explicarlas

debido a los cambios tecnológicos y de eficiencia que se irán produciendo.

Independientemente de la metodología matemática o econométrica elegida para la estimación final, el seguimiento sistemático de unidades productivas y el análisis de las diferencias en la productividad de las industrias a lo largo del tiempo debería ser un objetivo prioritario para institucionalizar la medición del desempeño de los servicios públicos.

2. Hacia una evaluación causal de los cambios en eficiencia y productividad

El siguiente paso, una vez lograda la institucionalización de la medición de los cambios en productividad de los distintos servicios públicos, será el análisis causal acerca de qué políticas públicas permiten mejorar el desempeño de las unidades productivas. Para ello, una línea actual de investigación consiste en combinar las herramientas tradicionales de inferencia causal con las técnicas de medición de la eficiencia basadas en el uso de fronteras. En este sentido, Mergoni y De Witte (2022) revisan 81 trabajos que realizan análisis causales en este sentido. De ellos, 53 trabajos tienen por objetivo analizar si cambios en políticas públicas o la aplicación de nuevos programas afectaron al desempeño de las unidades de producción respecto a un grupo de control. La mayoría de las aplicaciones reportadas son en el ámbito educativo [22], seguidas de servicios locales y ayuntamientos [13] y del sector salud [10], siendo el método de diferencias en diferencias el más aplicado (10).

Además de la combinación de fronteras productivas y técnicas de inferencia causal, una vez institucionalizada la medición del desempeño, en el futuro, también podrían diseñarse experimentos controlados aleatorizados para evaluar mejoras en productividad. En este sentido, Santín y Sicilia (2017) señalan cómo un nuevo programa público puede tardar bastante tiempo, debido a la necesidad de ganar experiencia o a problemas de gestión y ejecución que surgen en la aplicación de procedimientos novedosos, en producir diferencias en resultados estadísticamente significativos en la media en los *outcomes* deseados.

Sin embargo, estos autores confirman mediante un experimento de Monte-Carlo que es más fácil evaluar en el corto plazo si el programa está consiguiendo mejoras en productividad debido a causas tecnológicas, esto es, porque las mejores prácticas del grupo tratado empiezan a obtener mejores resultados que las mejores prácticas del grupo de control, o de eficiencia, cuando la eficiencia media de las unidades tratadas es significativa-

mente mejor que las del grupo de control. El gráfico 4 resume estas ideas.

El panel A del gráfico 4 muestra las unidades productivas de una industria que utiliza una única tecnología para producir una cantidad de *output* a partir de una cantidad de *input*. Mediante un proceso de aleatorización podemos dividir a las unidades en dos grupos; de tratamiento (T) y de control (C), representados en el gráfico por círculos blancos y negros respectivamente, cuyas medias en *inputs*, *outputs* y niveles de eficiencia no son estadísticamente distintas en el período base. El panel B del gráfico 4 muestra que si bien después de la aplicación del programa la eficiencia media de ambos grupos no es estadísticamente distinta $\bar{u}_T \cong \bar{u}_C$, la tecnología del grupo tratado domina ahora a la del grupo de control, indicando que el programa, si es bien implementado, puede mejorar la productividad de los servicios públicos. A partir de este análisis, y mediante una evaluación más detallada y cualitativa, se puede analizar las razones y canales que han motivado que estas unidades hayan mejo-

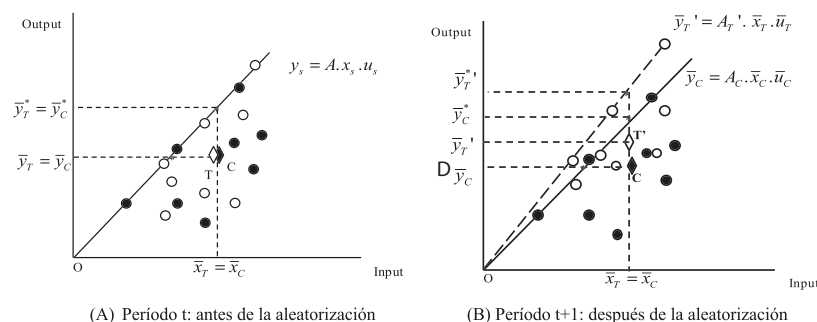
rado, aprender de su experiencia y trasladar estas buenas prácticas al resto del grupo tratado en períodos sucesivos.

3. Fusiones y escisiones de unidades productivas

Una decisión relevante en una institución es la escala con la que llevará a cabo sus operaciones. Pensemos, por ejemplo, en el caso de una universidad pública. Una universidad normalmente no opera de manera centralizada, lo habitual es que opere a partir de distintas facultades que, por ejemplo, agrupan distintas titulaciones relacionadas entre sí. Además, los profesores con intereses investigadores y/o docentes que pertenecen a una misma área o ámbito de conocimiento suelen agruparse por departamentos que a su vez pueden pertenecer a una o varias facultades. Pero, ¿cuál es el número óptimo de facultades y departamentos en una universidad (o en cualquier otra entidad pública que se organizara de una manera similar)?

La respuesta a esta pregunta es compleja ya que cada unidad de servicio supone distintos costes fijos y variables asociados a su proceso productivo. En este caso, al fusionar departamentos se podrían reducir los costes variables al liberar puestos de personal de administración y servicios de los departamentos fusionados para que realicen otras tareas o se pueden obtener ahorros por la reducción de complementos retributivos o de horas liberadas en gestión que ahora se podrían dedicar a tareas docentes e investigadoras. Por otro lado, tampoco parece sensato centralizar masivamente todos los servicios en un único departamento. Los costes de

GRÁFICO 4
ESQUEMA DE UN EXPERIMENTO CONTROLADO ALEATORIZADO UTILIZANDO FRONTERAS DE PRODUCCIÓN



Fuente: Elaboración propia a partir de Santín y Sicilia (2017).

tiempo en coordinación podrían ser tan elevados que merecería la pena dividir la unidad en varias unidades que operen de manera independiente. La casuística es elevada y, por tanto, es necesario sistematizar la búsqueda de las ganancias o pérdidas potenciales para cada caso.

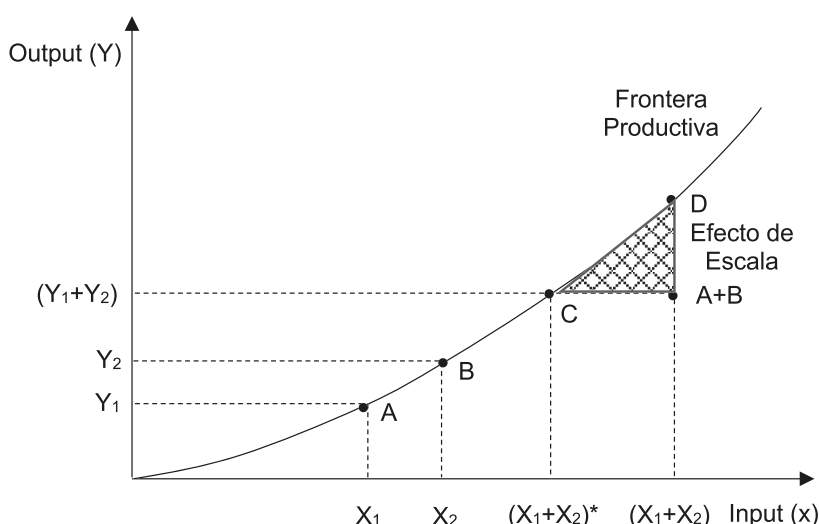
Las técnicas de fronteras productivas pueden ser utilizadas para analizar estas ganancias esperadas en eficiencia como consecuencia de fusionar o dividir unidades que gestionan el mismo servicio. En principio, en este ejemplo, una fusión entre varios departamentos será deseable si el nuevo departamento es capaz de lograr obtener una mayor producción en términos de horas de docencia y resultados de investigación o de alcanzar la misma producción a un menor coste respecto a los departamentos iniciales.

Siguiendo a Bogetoft y Wang (2006) y a Bogetoft y Otto (2011), las ganancias totales de eficiencia asociadas a la fusión pueden ser descompuestas en tres ganancias potenciales; efecto aprendizaje, efecto de armonización y efecto tamaño. El primer efecto es el denominado efecto aprendizaje (*learning effect*), o también llamado efecto eficiencia, y muestra cuánto puede mejorar la eficiencia del servicio si la nueva unidad opera en la frontera productiva. Obviamente, si las dos unidades son fusionadas «por decreto-ley», pero nada cambia, seguramente esta ganancia potencial será equivalente a la ganancia que las unidades previas alcanzarían si también alcanzaran la eficiencia. Sin embargo, la fusión podría buscar cambiar inercias negativas en la gestión que no cambiarían si no se introducen cambios (11).

La segunda ganancia potencial es la denominada efecto escala o efecto tamaño (*scale effect*). Este efecto se ilustra en el gráfico 5 donde se representa una tecnología en la que existen dos unidades productivas A y B que operan de manera plenamente eficiente al estar situadas sobre la frontera productiva. Si existen economías de escala, y por tanto con un tamaño mayor se pueden reducir los costes medios, la fusión de A+B permitiría llevar a cabo la misma cantidad de producción (Y_1+Y_2), pero reduciendo los inputs utilizados desde (X_1+X_2) hasta $(X_1+X_2)^*$, siendo $(X_1+X_2)^* < (X_1+X_2)$. En el gráfico 5 la ganancia potencial vendría dada por el triángulo (A+B)CD. Cabe destacar que esta ganancia podría ser negativa. El resultado final dependerá de la tecnología productiva que determinan las unidades productivas antes de la fusión, tecnología que deberá ser analizada mediante la construcción de una frontera productiva. De ahí la importancia de la evaluación.

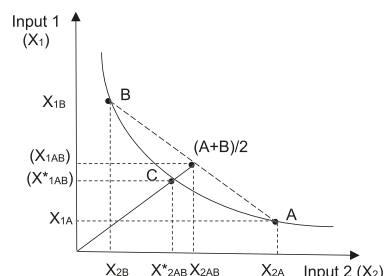
La tercera ganancia potencial es el efecto de armonización (*harmony effect*) que podemos ilustrar en el gráfico 6. En ella, se representa una isocuanta unitaria que reproduce la combinación de dos *inputs* capaces de producir una unidad de *output*. Se puede observar como las unidades A y B están muy desequilibradas en cuanto a la intensidad del uso de los factores productivos ya que A (B) utiliza muchas (pocas) unidades del *input* 1 y pocas (muchas) del *input* 2. En este caso, la fusión de ambas unidades genera una nueva unidad productiva, representada por el punto (A+B)/2. Tras la fusión vemos como la tecnología permite una ganancia de eficiencia potencial ya que la unidad fusionada podría reducir equiproportionalmente sus *inputs* hasta el punto C liberando estos recursos para usos alternativos y continuar produciendo la misma cantidad de producción.

GRÁFICO 5
EFECTO TAMAÑO O DE ESCALA EN LA FUSIÓN DE UNIDADES DE GESTIÓN



Fuente: Adaptado de Bogetoft y Otto (2011, p. 266)

GRÁFICO 6
EFFECTO ARMONIZACIÓN
EN LA FUSIÓN DE UNIDADES
DE GESTIÓN



Fuente: Adaptado de Bogetoft y Otto (2011, p. 270).

Las aplicaciones potenciales de esta metodología son numerosas tanto dentro de las organizaciones o servicios como entre ellos. Un ejemplo sería el de la fusión de municipios. España cuenta con más de 8.100 municipios y de ellos aproximadamente el 60 por 100 tiene menos de 1.000 habitantes. Salvo alguna excepción, como la fusión entre Oza dos Ríos y Cesuras, dos municipios de la provincia de A Coruña llevada a cabo en 2013, y la lograda entre las localidades de Don Benito y Villanueva de la Serena de la provincia de Badajoz en 2022, existen pocas iniciativas para la fusión de ayuntamientos.

Por ejemplo, para el caso de la integración de estos dos municipios pacenses, Cordero Pedraja-Chaparro y Suárez Pandiello (2022) señalan, entre otras numerosas ventajas, las ganancias en economías de escala; que incluye una posición más favorable para negociar con proveedores, reducción de costes medios y ganancias de financiación por aumento de población, al incrementarse del 1,3 al 1,5 el coeficiente multiplicador aplicado a los ingresos procedentes de

la participación en los ingresos tributarios del estado. Ganancias en eficiencia, al reducirse los costes en administración y mejorar la gestión de los recursos técnicos y humanos. Por ejemplo, la Ley de Régimen Electoral General concede 21 concejales para dos municipios entre 20.001 y 50.000 habitantes, de ellos entre 10 y 11 podrían tener dedicación exclusiva y percibir remuneración. Una eventual fusión por encima de 25.000 habitantes aumentaría a 25 el número de concejales y a 15 aquellos con dedicación a tiempo completo, lo cual supone un ahorro de recursos respecto a la suma de los dos municipios iniciales. Finalmente, la fusión permite ganancias de armonización, a la hora de alcanzar una mayor coherencia en la planificación económica del municipio aprovechando los recursos humanos e infraestructuras ya existentes, en lugar de proceder de manera individual a mejorar las carencias que tuviera cada municipio.

En definitiva, el análisis comparado de desempeño en el sector público permitiría cuantificar las ganancias y costes que se podrían obtener de potenciales fusiones para que una vez calculadas los ciudadanos y políticos decidan con mejor información si desean continuar adelante con la propuesta de fusión.

IV. CONCLUSIONES

La discusión realizada en las secciones precedentes pretende poner de manifiesto que:

– Siendo sumamente importante la evaluación de programas mediante inferencia causal basada en el uso de contrafactuales, la evaluación sistemática de la eficiencia y la productividad de

las unidades productoras de servicios públicos generaría fuertes ganancias de información para la toma de decisiones. Como ya se ha apuntado, un punto porcentual de mejora en la eficiencia de los servicios sanitarios y educativos supondría liberar 1.220 millones de euros para usos alternativos. Aunque en los últimos años se han dado pasos tímidos en el buen sentido, España está todavía lejos de alcanzar este objetivo.

– El primer requisito, en mi opinión, para llevar a cabo una adecuada evaluación sería crear agencias autonómicas independientes de evaluación de todos los programas y servicios públicos que dependan de los parlamentos autonómicos. En España, las CC. AA. tienen las competencias en sanidad y educación, los dos principales motores de gasto público «en especie». Por tanto, la administración autonómica, que es quien asigna el presupuesto, debería ser también la encargada en evaluar «sus propias industrias» (12). El objetivo inicial de cada agencia, que podría agrupar simultáneamente a varias CC. AA. si así lo deciden, sería disponer de información de calidad de manera estable en el tiempo para poder construir los indicadores necesarios para los análisis. Estos indicadores serían transparentes y estarían disponibles para que terceras partes pudieran realizar sus propios análisis. Este sistema de evaluación autonómica puede generar en el futuro problemas de coordinación e intercambio de información con agencias de otras CC. AA., pero este será un problema a medio plazo y de más fácil solución, al ya existir los mecanismos que generan la información.

– Construido el sistema anterior, el siguiente paso natural sería que varias CC. AA. se coordinen para comparar sus servicios utilizando los mismos indicadores. Ello pasará por una definición fina y común de los indicadores de consenso elegidos que permita la comparación. Con el tiempo, no entrar en el sistema interterritorial de evaluación tendrá un coste de reputación, como por ejemplo lo tendría salir de las evaluaciones internacionales PISA, TIMSS o PIRLS.

– Finalmente, levantado el sistema de información para cada servicio, las CC. AA. podrán explorar las ganancias (o pérdidas) en productividad que sus nuevas políticas tienen o tendrían en los servicios. Para ello se utilizarían, entre otras, las aproximaciones comentadas en la sección tercera. El indicador final de eficiencia junto a otras dimensiones, por ejemplo, de eficacia (mirando únicamente el lado de los *outputs*, de calidad o de reducción de *bad outputs*), desigualdad en la recepción de los servicios, etc., ofrecerá mayor transparencia para explicar las decisiones que afectan al servicio y al final los ciudadanos decidirán con sus votos la orientación final de las políticas.

Para terminar, no está mal resumir la importancia de evaluar utilizando la frase atribuida al célebre físico y matemático William Thomson Kelvin quien en el S. XIX afirmó: «Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre».

NOTAS

(1) A lo largo de este artículo, conceptos en economía como eficiencia, productividad, eficacia o evaluación del desempeño, se

entenderán en general referidos en relación con una frontera productiva o tecnología que subyace al proceso productivo y que debe ser estimada para permitir el *benchmarking* entre las unidades productivas comparadas.

(2) Abordar el detalle de todas y cada una de las últimas novedades metodológicas paramétricas y no paramétricas para medir eficiencia y productividad excede los objetivos de este artículo. Para comenzar con esta laboriosa tarea recomiendo el manual de SICKLES y ZELENYUK (2019).

(3) Para una excelente revisión del panorama actual y del marco regulador de la evaluación de políticas públicas en España se puede consultar DE LA FUENTE *et al.* (2021).

(4) No es el objetivo de este artículo realizar una revisión pormenorizada de todos los *inputs* y *outputs* que pueden ser utilizados en cada sector y de las metodologías que principalmente se usan para su evaluación. Para una revisión de variables y metodologías aplicadas en los principales sectores donde opera el sector público puede consultarse DE WITTE y LÓPEZ-TORRES (2017) para educación; HOLLINGSWORTH (2008) para el sector salud; NARBÓN-PERPIÑÁ y DE WITTE (2018a y 2018b) para servicios locales y ayuntamientos, y ESPASA y ESTELLÉ-MORÉ (2015) para juzgados de primera instancia.

(5) Además de los clásicos *inputs* y *outputs*, el análisis de eficiencia permite la inclusión de los denominados *bad outputs* para incorporar indicadores no deseables asociados a la producción que deberían ser también tenidos en cuenta para evaluar el desempeño. Para una revisión de los mismos aplicada al contexto de comparación de sistemas educativos puede acudirse a GIMÉNEZ *et al.* (2017).

(6) Para un estudio acerca de las principales metodologías disponibles para introducir variables de entorno en los análisis de eficiencia puede acudirse a DARAIO y SIMAR (2007); JOHNSON y KOUSMANEN (2011) y SIMAR y WILSON (2011).

(7) A lo largo del artículo utilizo el término genérico de «industria» para referirme a cualquier servicio público que cumpla las características de esta definición.

(8) En Fórmula 1 esto sería equivalente a comprobar que cada año nuestro coche, dando vueltas al mismo trazado, reduce los tiempos. Aunque, sin duda, este cambio parece ir por el buen sentido, esta evaluación resulta vacía sin comparaciones. El siguiente paso sería poner a competir a dos coches de nuestro mismo equipo que nos sirva para incentivar una sana competencia entre ingenieros y pilotos con ánimo de introducir mejoras en los coches. Finalmente, tendremos que compararnos con coches de otros equipos.

(9) Una limitación de esta metodología es que las medidas de diferencias en producti-

vidad pueden variar al cambiar la referencia utilizada. Sin embargo, APARICIO, ORTIZ y SANTÍN (2021) muestran cómo la correlación entre las medidas obtenidas utilizando dos referencias distintas varía entre 0,94 y 0,99. Como ya se ha señalado, idealmente, evaluadores, unidades evaluadas y los usuarios de la evaluación deberían decidir qué tecnología usar como referencia que incluso podría ser la metafrontera en un período base.

(10) Por ejemplo, FELICIANO, LÓPEZ-TORRES y SANTÍN (2021) evalúan el programa Escuela 2.0 comparando la evolución de las diferencias en productividad de los resultados en Educación Secundaria de las CC. AA. que aplicaron el programa respecto a la Comunidad de Madrid que no lo aplicó. Los principales resultados muestran que las CC. AA. que aplicaron el programa perdieron en promedio un 13,3 por 100 de productividad respecto a Madrid.

(11) En cualquier caso, dentro de esta primera ganancia de eficiencia es fácil cuantificar la eliminación de costes duplicados que automáticamente pueden ser reducidos. En el ejemplo hipotético de la fusión de dos departamentos de universidad pequeños, digamos de 20 profesores que pasan a ser un único departamento de 40 profesores, se pasaría de dos directores, dos secretarios académicos y dos personas de administración y servicios a solo una, pudiendo dedicar los recursos liberados a otras tareas.

(12) En mi opinión, una única Agencia Nacional como la que plantean DE LA FUENTE *et al.* (2021) que lleve la evaluación «de arriba hacia abajo» va a encontrar mayores problemas de coordinación y reticencias a la hora de obtener toda la información que si se opta por agencias autonómicas que realicen la evaluación «de abajo a arriba». Algunas de estas agencias podrían nacer por acuerdo para evaluar a varias comunidades autónomas simultáneamente. Con la escasa cultura evaluadora existente en España, una Agencia Nacional será percibida como un organismo fiscalizador y habrá reticencias y retrasos a la hora de compartir con ella la información. Sin embargo, una agencia propia puede contribuir a obtener mejor información y a incidir en que la evaluación solo tiene como objetivo proporcionar información veraz para mejorar los servicios.

BIBLIOGRAFÍA

AIREF (2017). *I Plan de Acción de la revisión del gasto en subvenciones del conjunto de las Administraciones Públicas. (Spending Review)*.

AIREF (2018). *Informe de evaluación de la Sociedad Estatal Correos y Telégrafos, y la prestación del servicio postal universal*.

- ÁLVAREZ, A. (Coord.) (2001). *La medición de la eficiencia y la productividad*. Madrid: Ediciones Pirámide.
- APARICIO, J., ORTIZ, L. y SANTÍN, D. (2021). Comparing group performance over time through the Luenberger productivity indicator: An application to school ownership in European countries. *European Journal of Operational Research*, 294(2), pp. 651-672.
- APARICIO, J. y SANTÍN, D. (2018). A note on measuring group performance over time with pseudo-panels. *European Journal of Operational Research*, 267(1), pp. 227-235.
- BATTESE, G. E. y RAO, D. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, 1(2), pp. 87-93.
- BERG, S. A., FØRSUND, F. R. y JANSEN, E. S. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics*, S211-S228.
- BOGETOFT, P. y OTTO, L. (2010). *Benchmarking with DEA, SFA, and R* (Vol. 157). Springer Science & Business Media.
- BOGETOFT, P. y WANG, D. (2005). Estimating the potential gains from mergers. *Journal of Productivity Analysis*, 23(2), pp. 145-171.
- CAMANHO, A. S. y DYSON, R. G. (2006). Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis*, 26(1), pp. 35-49.
- CAVES, D., CHRISTENSEN, L. y DIEWERT, W. (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica*, 50(6), pp. 1393-1414.
- COELLI, T. J., RAO, D. S. P., O'DONNELL, C. J. y BATTESE, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.
- COLEGIO DE ECONOMISTAS DE MADRID (2005). La eficiencia de los servicios públicos: viejos problemas, nuevos enfoques. *Revista Economistas*, 105. Madrid: Colegio de Economistas de Madrid.
- CORDERO, J. M., PEDRAJA-CHAPARRO, F. M. y SUÁREZ PANDIELLO, J. (2022). *Estudio sobre organización territorial municipal en la provincia de Badajoz*. Mimeo. Recuperado de: <https://www.donbenito.es/wp-content/uploads/2022/01/INFORME-FINAL-CON-RESUMEN-EJECUTIVO-ESTUDIO-DE-ORGANIZACION-TERRITORIAL-PEDRAJA-PANDIELLO.pdf>
- CORDERO, J. M., SANTÍN, D. y SIMANCAS, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of Operational Research Society*, 68(4), pp. 364-376.
- DARAIO, C. y SIMAR, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, 28(1), pp.13-32.
- DE LA FUENTE, Á. (Coord.) (2021). La evaluación de políticas públicas en España: antecedentes, situación actual y propuestas para una reforma. *Fedea Policy Papers*, 2021-09. Fedea.
- ESPASA, M. y ESTELLER-MORE, A. (2015). Analyzing judicial courts' performance: inefficiency vs. congestion. *Revista de Economía Aplicada*, 23(69), pp. 61-82.
- FARRELL, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, 120(3), pp. 253-281.
- FELICIANO, D., LÓPEZ-TORRES, L. y SANTÍN, D. (2021). One Laptop per Child? Using Production Frontiers for Evaluating the Escuela 2.0 Program in Spain. *Mathematics*, 9(20), 2600.
- FERREIRA, D. y MARQUÉS, R. C. (2015). Did the corporatization of Portuguese hospitals significantly change their productivity? *The European Journal of Health Economics*, 16(3), pp. 289-303.
- FERREIRA, D. C., MARQUES, R. C. y PEDRO, M. I. (2016). Comparing efficiency of holding business model and individual management model of airports. *Journal of Air Transport Management*, 57, pp.168-183.
- FRIED, H. O., LOVELL, C. K. y SCHMIDT, S. S. (Eds.). (2008). *The measurement of productive efficiency and productivity growth*. Oxford University Press.
- FUNCAS (2003). Sector público y eficiencia. *Papeles de Economía Española*, 95. Madrid: Funcas.
- GIMÉNEZ, V., THIEME, C., PRIOR, D. y TORTOSA-AUSINA, E. (2017). An international comparison of educational systems: a temporal analysis in presence of bad outputs. *Journal of Productivity Analysis*, 47(1), pp. 83-101.
- GIMÉNEZ, V., THIEME, C., PRIOR, D. y TORTOSA-AUSINA, E. (2019). Comparing the performance of national educational systems: inequality versus achievement? *Social Indicators Research*, 141(2), pp. 581-609.
- JOHNSON, A. L. y KUOSMANEN, T. (2011). One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNEZD method. *Journal of Productivity Analysis*, 36(2), pp. 219-230.
- KUMBHAKAR, S. C., WANG, H. y HORNCastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.
- LEIBENSTEIN, H. (1966). Allocative efficiency vs. «X-efficiency». *The American Economic Review*, 56(3), pp. 392-415.
- LEVIN, H. M. (1974). Measuring efficiency in educational production. *Public Finance Quarterly*, 2(1), pp. 3-24.
- LEVIN, H. M., JAMISON, D. T. y RADNER, R. (1976). Concepts of economic efficiency and educational production. *Education as an Industry*, pp. 149-198. NBER.
- MAS-COLELL, A., WHINSTON, M. D. y GREEN, J. R. (1995). *Microeconomic theory* (Vol. 1). New York: Oxford University Press.
- MERGONI, A. y DE WITTE, K. (2022). Policy evaluation and efficiency: a systematic literature

review. <i>International Transactions in Operational Research</i>, 29(3), pp. 1337-1359. MEFP (2021). <i>Anuario estadístico: Las cifras de la educación en España. Edición 2021</i> (Curso 2018-2019). Madrid: Ministerio de Educación y Formación Profesional. MINISTERIO DE SANIDAD (2021). <i>Informe Anual del Sistema Nacional de Salud 2019. Aspectos destacados</i>. Madrid: Ministerio de Sanidad. NARBÓN-PERPIÑÁ, I. y DE WITTE, K. (2018a). Local governments» efficiency: a systematic literature review – part I. <i>International Transactions in Operational Research</i>, 25(2), pp. 431-468. NARBÓN-PERPIÑÁ, I. y DE WITTE, K. (2018b). Local governments» efficiency: a systematic literature review – part II. <i>International Transactions in Operational Research</i>, 25(4), pp. 1107-1136.	NISKANEN, W. A. (1975). Bureaucrats and politicians. <i>The Journal of Law and Economics</i>, 18(3), pp. 617-643. OATES, W. E. (1999). An essay on fiscal federalism. <i>Journal of Economic Literature</i>, 37(3), pp. 1120-1149. OH, D. H. (2010). A global Malmquist-Luenberger productivity index. <i>Journal of Productivity Analysis</i>, 34(3), pp. 183-197. PASTOR, J. T. y LOVELL, C. K. (2005). A global Malmquist productivity index. <i>Economics Letters</i>, 88(2), pp. 266-271. PEDRAJA-CHAPARRO, F. M., SALINAS-JIMÉNEZ, J. y SMITH, P. (1999). On the quality of the data envelopment analysis model. <i>Journal of the Operational Research Society</i>, 50(6), pp. 636-644. SANTÍN, D. Y SICILIA, G. (2017). Impact evaluation and frontier methods in education: a step forward. En:	G. JOHNES, J. JOHNES, T. AGASISTI y L. LÓPEZ-TORRES (eds.), <i>Handbook of Contemporary Education Economics</i>. Cheltenham: Edward Elgar Publishing. SCRGSP, <i>Steering Committee for the Review of Government Service Provision</i> (2013). <i>Report on Government Services 2013</i>. Canberra: Productivity Commission. SICKLES, R. C. y ZELENYUK, V. (2019). <i>Measurement of productivity and efficiency</i>. Cambridge University Press. SIMAR, L. y WILSON, P. W. (2011). Two-stage DEA: caveat emptor. <i>Journal of Productivity Analysis</i>, 36(2), pp. 205-218. TIROLE, J. (1994). The internal organization of government. <i>Oxford Economic Papers</i>, 46, pp. 1-29.
---	--	--

COLABORACIONES

II. IMPUESTOS Y POLÍTICA FISCAL

Resumen

Utilizando una encuesta *online* llevada a cabo en mayo de 2020, encuentro que los individuos en hogares de más renta expresan una mayor preferencia por subir los impuestos y son más proclives a expresar una opinión positiva sobre ellos. Esta correlación se mantiene incluso al controlar por una serie de importantes factores sociodemográficos y políticos. Además, los individuos varones, de más edad, mayor nivel educativo, votantes de izquierdas y, muy especialmente, aquellos con más confianza en el sistema político expresan una mayor preferencia por subir los impuestos.

Palabras clave: instituciones, confianza, impuestos, economía política, COVID-19.

Abstract

Using an online survey carried out in May 2020, I find that individuals in higher income households express a greater preference for raising taxes and are more likely to express a positive opinion about them. This correlation holds even when controlling for a number of important sociodemographic and political factors. In addition, male, older, more educated, and left-leaning individuals, and, especially, those with more trust in the political system, express a greater preference for raising taxes.

Keywords: institutions, trust, taxes, political economy, COVID-19.

JEL classification: D70, H10, H20, H30, H50, H60.

¿MÁS O MENOS IMPUESTOS? RENTA, CONFIANZA EN LAS INSTITUCIONES Y PREFERENCIAS IMPOSITIVAS

Carlos SANZ (*)

Banco de España

I. INTRODUCCIÓN

¿QUÉ factores determinan el tamaño del sector público de un país? ¿Cuáles son las preferencias de los ciudadanos a este respecto? Estas preguntas han atraído la atención de la literatura económica desde hace décadas y han cobrado especial relevancia tras la pandemia de la COVID-19 y el consiguiente aumento de los déficits públicos. Por ejemplo, según datos de la Intervención General de la Administración del Estado, el gasto público en España en 2020 alcanzó el 52,3 por 100 del PIB, una subida de 10,2 puntos porcentuales respecto a 2019.

En este artículo, se aporta nueva evidencia sobre las preferencias de la población española sobre el tamaño del sector público, basándose en los datos obtenidos en una encuesta *online* llevada a cabo con Mónica Martínez-Bravo (CEMFI) en colaboración con la empresa de datos YouGov (1). Más en concreto, se estudian las preferencias de los individuos sobre los impuestos y qué factores económicos, sociodemográficos y políticos explican dichas preferencias.

En primer lugar, y de manera un tanto sorprendente, se documenta que los individuos que viven en hogares de más

renta expresan una mayor preferencia por subir los impuestos y son más proclives a expresar una opinión positiva sobre ellos. En la medida en que puede esperarse que mayores impuestos se asocien a más redistribución, este hallazgo es contrario a los supuestos teóricos de los modelos clásicos de economía política (Meltzer y Richard, 1981). En segundo lugar, se estudia qué otros factores sociodemográficos y políticos correlacionan con las preferencias impositivas de los ciudadanos. Encontramos que los hombres y las personas de más edad y más nivel educativo, así como los votantes de izquierdas (PSOE y UP) expresan una mayor preferencia por subir los impuestos. Las diferencias por nivel de ingresos en las preferencias impositivas, no obstante, se mantienen aun después de tener en cuenta todos estos factores. Por último, nos encontramos que una variable clave para explicar las preferencias impositivas es la confianza en las instituciones públicas y el sistema político: las personas con más confianza expresan una mayor preferencia por subir los impuestos y son más proclives a mencionar opiniones positivas sobre los mismos. Esta correlación es cuantitativamente importante, estadísticamente significativa y está presente después de controlar por el resto de factores de renta, sociodemográficos y políticos.

Este artículo contribuye a la literatura que estudia los determinantes del tamaño del sector público. El clásico modelo del votante mediano predice que una mayor desigualdad debería llevar a más impuestos y mayor redistribución (Meltzer y Richard, 1981). Intuitivamente, dado que el votante mediano es más pobre (antes de impuestos) que la media, esto dará lugar a más impuestos y redistribución. Aumentos de la desigualdad (normalmente asociados a una mayor diferencia entre la media y la mediana) conducirán de este modo a un mayor nivel de impuestos. Un supuesto clave de estos modelos es que los individuos votan de acuerdo con su propio interés y que este depende exclusivamente de la renta. De este modo, el modelo supone que los individuos más ricos votarán en favor de menores impuestos que los pobres, contrariamente a la evidencia que el presente artículo presenta para España (2).

La literatura posterior a Meltzer y Richard (1981) ha intentado analizar la relación predicha por el modelo de que mayor desigualdad antes de impuestos debería llevar a mayor redistribución. Alesina y Glaeser (2004) encuentran una correlación negativa entre desigualdad de mercado y redistribución, en un análisis de diferencias entre países desarrollados. Kenworthy y McCall (2008) estudian la evolución temporal dentro de ocho países de la OCDE y tampoco encuentran apoyo para el modelo de Meltzer y Richard (1981). Una posible explicación podría ser que, aunque no se observe una correlación positiva entre desigualdad de mercado y redistribución, esto sea debido a un fallo del sistema político y

no a la preferencia de los votantes. Sin embargo, Kenworthy y McCall (2008) tampoco encuentran una correlación entre aumentos de la desigualdad y las preferencias por redistribución declaradas por los ciudadanos en encuestas.

Estos resultados no implican que el mecanismo de Meltzer y Richard (1981) no esté en juego, pero sí que debe haber otros factores más importantes que hagan que la correlación no sea la esperada. Por este motivo, más recientemente la literatura ha propuesto nuevas teorías sobre los determinantes de la redistribución (ver Alesina y Giuliano, 2011, para una revisión de la literatura). En este sentido, se ha encontrado que el nivel de redistribución depende de las creencias de los individuos sobre la justicia de la distribución de rentas de mercado. Teóricamente, esta posibilidad ha sido desarrollada por Piketty (1995) y Benabou y Tirole (2006), entre otros. Empíricamente, Alesina y La Ferrara (2005), Alesina y Giuliano (2011), y Fong (2001) encuentran, respectivamente, que las creencias sobre la movilidad de la renta en el futuro, experiencias pasadas de mala fortuna y las creencias sobre la igualdad de oportunidades predicen las preferencias sobre redistribución. Finalmente, y muy relacionado con el presente artículo, hay evidencia de que la confianza de los individuos en las herramientas redistributivas de la política económica influye en las preferencias sobre redistribución. En este sentido, Kuziemko *et al.* (2015) encuentran que dar información detallada sobre la desigualdad de renta afecta significativamente a la preocupación por la desigualdad, pero apenas afecta a

las preferencias sobre impuestos y transferencias (con la notable excepción del impuesto de sucesiones o *estate tax*). Encuentran que esto se debe a la falta de confianza en el Gobierno y en la desconexión percibida entre los problemas sociales y las políticas públicas dirigidas a abordarlas (3).

El presente artículo contribuye también a la literatura que estudia la relevancia de la confianza en las instituciones. Existe considerable evidencia de que la confianza es crucial para el desarrollo económico y el buen funcionamiento institucional (ver, por ejemplo, Algan y Cahuc [2010], y Acemoglu *et al.* [2020]; ver también la revisión de esta literatura de Sorribas-Navarro [2021], en este mismo monográfico). Los resultados del presente artículo apuntan también a una fuerte correlación de la confianza en el sistema político y las instituciones con las preferencias con el tamaño del sector público (4).

El enfoque del artículo es fundamentalmente descriptivo; es decir, describe las variables de interés y estudia qué relaciones tienen entre sí, aunque sin identificar efectos causales. El análisis descriptivo es una etapa necesaria en cualquier análisis científico, ya que facilita la comprensión de las asociaciones existentes entre las distintas variables y sirve como motivación y generador de ideas para posteriores análisis causales.

El resto del artículo se organiza de la siguiente manera. La sección segunda describe la encuesta y los datos utilizados. La tercera sección presenta los resultados. La sección cuarta concluye el artículo, sugiriendo nuevas líneas de investigación.

II. DATOS

1. Descripción de la encuesta

Los datos analizados en este trabajo se han obtenido a partir de dos olas de una encuesta *online* realizada durante los meses de mayo y noviembre de 2020. Se entrevistó a personas mayores de 18 años residentes en España. El trabajo de campo fue llevado a cabo por la empresa YouGov, que dispone de un panel de conveniencia de más de 50.000 personas, a las que se da un incentivo monetario para contestar una serie de encuestas. El muestreo de la primera ola se diseñó para tener una muestra que replique la estructura de la población española por edad, género, región (área Nielsen) y nivel educativo (5). Todos los individuos que completaron la primera ola volvieron a ser contactados para la segunda, respondiendo a la segunda un 65 por 100 de los inicialmente entrevistados. Dado que no se encuentran apenas diferencias entre los resultados de la primera ola y la segunda, el presente artículo se centra en la primera ola (6). En concreto, la muestra utilizada comprende 4.566 personas que completaron el cuestionario de la primera ola y proporcionaron información de todas las variables utilizadas (descritas en la siguiente sección). YouGov proporciona pesos obtenidos por el *Random Iterative Method* que calibran a las distribuciones marginales de edad, género, región y nivel educativo de la población. Se utilizan estos pesos en todos los análisis del artículo. Los resultados son muy similares si los pesos no se utilizan (7).

2. Variables y descriptivos

En esta sección se describen los datos utilizados en el presente trabajo.

Variables dependientes. Las variables dependientes se construyen a partir de dos preguntas. En la primera de ellas, se pide a los encuestados que indiquen su preferencia por los impuestos en una escala del 0 al 10, donde 0 es «bajar impuestos» y 10 «subir impuestos» (8). El cuadro n.º 1, que muestra los estadísticos descriptivos de las variables utilizadas en el presente artículo, indica que los individuos se colocan, de media, en un 3.90.

En la segunda pregunta, se pide a los encuestados indicar cuál o cuáles de las siguientes frases reflejan mejor su opinión sobre los impuestos, permitiéndose seleccionar más de una respuesta de entre las siguientes:

1. «Los impuestos son un medio para redistribuir mejor la riqueza en la sociedad».

2. «Los impuestos son necesarios para que el Estado pueda prestar servicios públicos».

3. «Lo que pagamos en impuestos no se corresponde con los servicios públicos que recibimos, por culpa de la corrupción».

CUADRO N.º 1

ESTADÍSTICOS DESCRIPTIVOS

	MEDIA	MÍN	MÁX	SD	CONTADOR
Preferencias impositivas					
Subir impuestos (0-10)	3,90	0,00	10,00	2,74	4.566
Necesarios redistribucion	0,21	0,00	1,00	0,41	4.566
Necesarios bienes públicos	0,49	0,00	1,00	0,50	4.566
Despilfarro	0,56	0,00	1,00	0,50	4.566
Mejor en manos privadas	0,17	0,00	1,00	0,38	4.566
Opinión positiva	0,35	0,00	1,00	0,48	4.566
Opinión negativa	0,44	0,00	1,00	0,50	4.566
Opinión mixta	0,21	0,00	1,00	0,41	4.566
Variables sociodemográficas					
Ingresos	2.599,28	0,00	24.000,00	2.152,69	4.566
Mujer	0,51	0,00	1,00	0,50	4.566
Edad <31	0,16	0,00	1,00	0,37	4.566
Edad [31,65]	0,74	0,00	1,00	0,44	4.566
Edad >65	0,10	0,00	1,00	0,30	4.566
Educ, Primaria o menos	0,17	0,00	1,00	0,38	4.566
Educ, Secundaria	0,28	0,00	1,00	0,45	4.566
Educ, Bachillerato	0,14	0,00	1,00	0,35	4.566
Educ, FP	0,08	0,00	1,00	0,27	4.566
Educ, Superior	0,32	0,00	1,00	0,47	4.566
Variables sociodemográficas					
Voto PSOE	0,19	0,00	1,00	0,39	4.566
Voto PP	0,09	0,00	1,00	0,29	4.566
Voto VOX	0,09	0,00	1,00	0,28	4.566
Voto UP	0,13	0,00	1,00	0,34	4.566
Voto Cs	0,05	0,00	1,00	0,22	4.566
Voto otro	0,44	0,00	1,00	0,50	4.566
Confianza (0-10)	4,16	0,00	10,00	2,92	4.566

Fuente: Estadísticos calculados usando pesos muestrales.

4. «El dinero que el Estado recauda con los impuestos estaría mejor en los bolsillos de los ciudadanos».

El objetivo de esta pregunta es inferir las actitudes positivas o negativas de los encuestados hacia los impuestos. A partir de ella, se generan siete variables. Las cuatro primeras, *Necesarios Redistribución*, *Necesarios Bienes Públicos*, *Despilfarro* y *Mejor en Manos Privadas*, son variables dicotómicas que toman el valor 1 si el encuestado selecciona la opción correspondiente y 0 si no lo hace. Las tres últimas, *Opinión Positiva*, *Opinión Negativa* y *Opinión Mixta* son variables dicotómicas que agregan la información de las cuatro anteriores: *Opinión Positiva* toma el valor 1 si el encuestado selecciona al menos una de las respuestas positivas (es decir, las opciones 1 o 2) y no selecciona ninguna de las respuestas negativas (es decir, ni la opción 3 ni la 4), y 0 en caso contrario. *Opinión Negativa* toma el valor 1 si el encuestado selecciona al menos una de las respuestas negativas sobre los impuestos y no selecciona ninguna de las respuestas positivas, y 0 en caso contrario. *Opinión Mixta* toma el valor 1 en el resto de los casos; es decir, si el encuestado selecciona opciones tanto positivas como negativas.

Como se desprende del cuadro n.º 1, las respuestas que más eligen los encuestados son la 3 (lo que pagamos no se corresponde con los servicios públicos), que es elegida por el 56 por 100 de los encuestados, y la 2 (los impuestos son necesarios para proveer servicios públicos), que es elegida por el 49 por 100 de los encuestados. Por debajo quedan la opción 1 (los impuestos son necesarios para redistribuir

renta) y la 4 (el dinero recaudado estaría mejor en manos de los ciudadanos), que son seleccionadas por el 21 por 100 y el 17 por 100, respectivamente.

Variables explicativas. Como variables explicativas, considero, en primer lugar, los ingresos mensuales medios per cápita del hogar en el que reside el encuestado en 2019 (9). En media, los encuestados viven en un hogar que obtiene unos ingresos mensuales de 2.599 euros per cápita, como puede verse en el cuadro n.º 1. En el análisis empírico, divido esta variable por quintiles en cinco variables dicotómicas.

En segundo lugar, se consideran variables sociodemográficas de género, edad (dividida en tres categorías: menores de 31 años, entre 31 y 65 años y mayores de 65 años) y nivel educativo (en cinco categorías: primaria o no disponible; secundaria; bachillerato; FP y superior).

En tercer lugar, se considera, como variables políticas, la intención de voto a los cinco grandes partidos: PSOE, PP, VOX, UP y Cs. La variable *Voto Otros* incluye votar tanto al resto de opciones como la abstención como el preferir no contestar a la pregunta. El cuadro n.º 1 muestra que el voto al PP está probablemente infrarrepresentado, lo que es común en las encuestas de intención de voto (Cabrera-Álvarez y Escobar, 2019). Los resultados del presente artículo son, en todo caso, similares si se utilizan variables de simpatía en lugar de intención de voto directa. También se toma en cuenta la autoubicación ideológica en una escala del 0 (izquierda) al 10 (derecha). Sin embargo, esta variable no es significativa ni afecta a los resultados una vez se in-

cluyen las variables de voto, por lo que se excluye de los resultados que se presentan.

Finalmente, se estima la confianza en el sistema político. La variable *Confianza* se genera a partir de la siguiente pregunta: «En una escala del 0 al 10, donde 0 es muy baja y 10 es muy alta, ¿cómo valorarías la capacidad del sistema político de proporcionar soluciones a los principales problemas de los ciudadanos?». Otras preguntas sobre confianza en instituciones concretas (confianza en el Gobierno central, en los Gobiernos autonómicos o en los diputados) dan lugar a resultados muy similares.

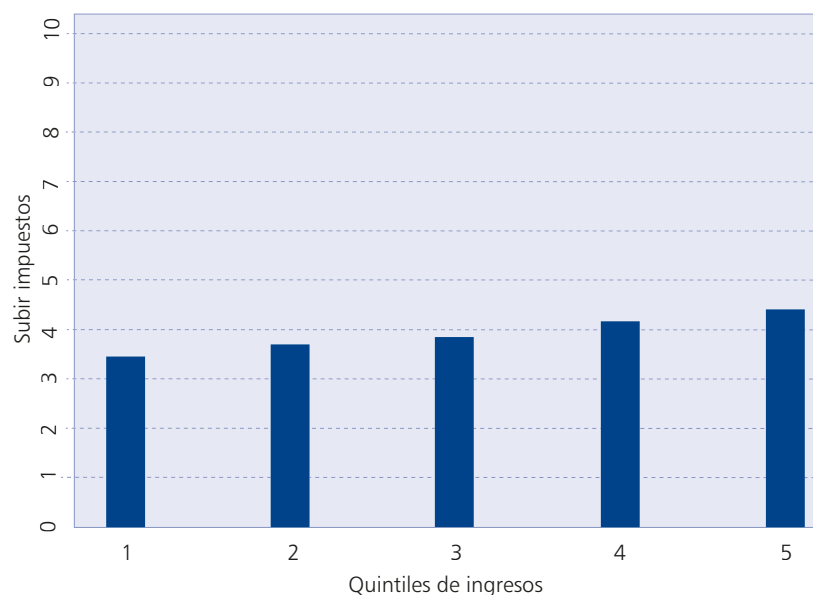
III. RESULTADOS

1. Impuestos e ingresos: análisis gráfico

Comenzamos viendo la relación entre ingresos y preferencias impositivas. El gráfico 1 muestra, para cada quintil de ingresos del hogar, la preferencia media sobre subir o bajar impuestos, donde 0 es bajarlos y 10 es subirlos. Como puede apreciarse, existe un gradiente positivo por ingresos: los individuos residentes en hogares más ricos tienen una mayor preferencia por subir impuestos. En la sección 3.2 se estudiará la significatividad estadística y las magnitudes en el marco de una regresión de mínimos cuadrados.

El gráfico 2 presenta las respuestas dadas por los encuestados cuando se les pregunta por su opinión sobre los impuestos. Se observa un gradiente por renta en tres de las cuatro respuestas: los individuos en hogares más ricos son más propensos a mencionar que los impuestos

GRÁFICO 1
PREFERENCIA POR SUBIR LOS IMPUESTOS, POR INGRESOS



Notas: Las barras muestran la preferencia media, del 0 al 10, sobre subir impuestos (10) o bajarlos (0). Los encuestados se agrupan en cinco quintiles de acuerdo con sus ingresos familiares per cápita en 2019. El quintil 1 es el quintil más pobre y el 5, el más rico.

Fuente: Elaboración propia.

son un medio para redistribuir mejor la riqueza en la sociedad (gráfico 2a) y necesarios para que el Estado pueda prestar servicios públicos (gráfico 2b), y menos propensos a contestar que el dinero que el Estado recauda con los impuestos estaría mejor en los bolsillos de los ciudadanos (gráfico 2d). No se observa un claro gradiente por ingresos en la propensión a contestar que lo que pagamos en impuestos no se corresponde con los servicios públicos que recibimos (gráfico 2c).

El gráfico 3 muestra los resultados tras agregar las cuatro respuestas anteriores. El gráfico 3a muestra que la probabilidad de dar una opinión positiva sobre los impuestos (y ninguna negativa) aumenta con los ingresos. El gráfico 3b muestra que la probabilidad de dar una opinión

negativa sobre los impuestos (y ninguna positiva) disminuye con los ingresos. Esto sugiere que los individuos más ricos tienen una visión más positiva de los impuestos que los más pobres.

2. Impuestos y otras características: análisis de regresión

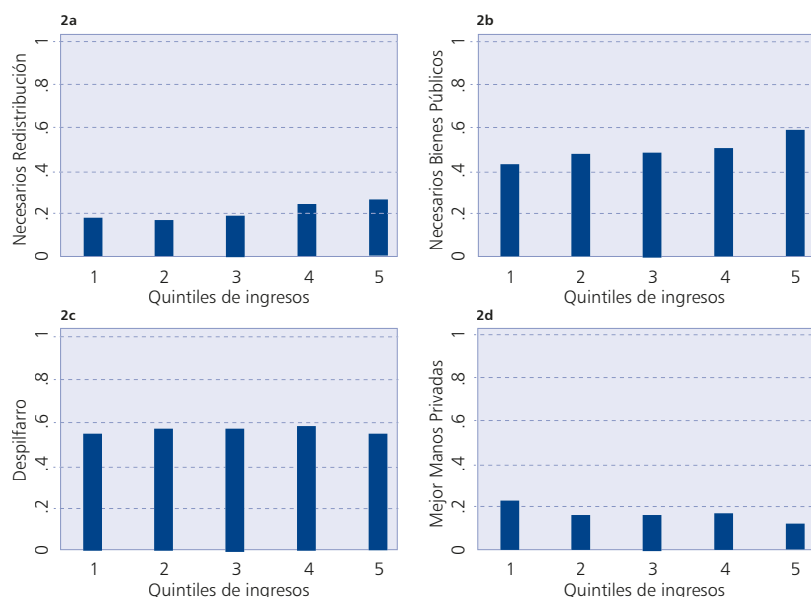
A continuación, se analiza en un marco de regresión múltiple la correlación entre las opiniones hacia los impuestos y un conjunto de variables sociodemográficas, políticas y de confianza en las instituciones, además de las variables de ingresos exploradas en la subsección anterior. El objetivo de este análisis es doble. En primer lugar, analizar si la correlación positiva entre ingresos y actitudes hacia los impuestos se

mantiene al controlar por otros factores relevantes. En segundo lugar, determinar qué otros factores, además de los ingresos, están correlacionados con las actitudes hacia los impuestos.

La primera columna del cuadro n.º 2 muestra los resultados de la regresión de *Subir Impuestos* sobre los cinco quintiles de ingresos. Es decir, los coeficientes se corresponden con las barras representadas en el gráfico 1. Los resultados indican que la preferencia por subir impuestos aumenta por quintil de forma aproximadamente lineal. Los coeficientes son cuantitativamente grandes: por ejemplo, los individuos en el quintil más rico dan una preferencia sobre impuestos casi un punto (sobre diez) más alta que los individuos en el quintil más pobre. Las diferencias son estadísticamente significativas para los quintiles 3, 4 y 5.

La segunda columna añade las variables sociodemográficas. Las mujeres tienen una preferencia significativamente menor por subir impuestos que los hombres: 0,26 puntos (en una escala 0-10). Respecto a los jóvenes (<31 años), los individuos de edad media (de 31 a 65 años) no muestran una diferencia significativa, mientras que los mayores (>65 años) tienen una mayor preferencia por aumentar los impuestos. La diferencia es cuantitativamente relevante: 0,57 puntos (en la escala 0-10). También hay diferencias significativas por nivel educativo: los resultados indican que los individuos con menos educación (menos de secundaria) tienen una menor preferencia por subir los impuestos. Por ejemplo, en relación con este grupo, las personas con educación superior dan una preferen-

GRÁFICO 2
OPINIONES SOBRE LOS IMPUESTOS, POR INGRESOS



Notas: Las barras muestran la proporción de encuestados que contestan «Los impuestos son un medio para redistribuir mejor la riqueza en la sociedad» (gráfico 2a), «Los impuestos son necesarios para que el Estado pueda prestar servicios públicos» (gráfico 2b), «Lo que pagamos en impuestos no se corresponde con los servicios públicos que recibimos, por culpa de la corrupción» (gráfico 2c), «El dinero que el Estado recauda con los impuestos estaría mejor en los bolsillos de los ciudadanos» (gráfico 2d). Los encuestados se agrupan en cinco quintiles de acuerdo con sus ingresos familiares per cápita en 2019. El quintil 1 es el quintil más pobre y el 5, el más rico.

Fuente: Elaboración propia.

cia por 0,46 puntos mayor por subir impuestos. No obstante, aunque estadísticamente significativas y cuantitativamente importantes, estas variables solo explican alrededor del 2 por 100 de la variación en las preferencias impositivas ($R^2=0,02$). Las variables de ingresos se reducen algo al controlar por las variables sociodemográficas, pero siguen siendo significativas.

También merece la pena señalar que controlar por la comunidad autónoma de residencia mediante efectos fijos (resultados no mostrados) no afecta ninguna de las correlaciones anteriores. Asimismo, tampoco se observan grandes diferencias

entre comunidades autónomas una vez se ha controlado por el resto de factores, aunque las estimaciones de estas diferencias no son muy precisas debido a la falta de tamaño muestral.

La tercera columna añade las variables de intención de voto. Respecto a la variable omitida, que es el voto al PSOE, los votantes de UP tienen una preferencia 0,63 puntos (sobre 10) mayor por subir los impuestos, mientras que los de Cs, PP y VOX tienen una preferencia 1,24, 2,17 y 2,79 puntos menor, respectivamente (10). Estas variables son cuantitativa y estadísticamente significativas y además tienen un considerable valor explicati-

vo: hacen subir el R^2 de ,02 a ,16. Las variables de ingresos y sociodemográficas se mantienen similares tras incluirse los controles de voto.

Por último, la cuarta columna introduce la confianza en el sistema político. El coeficiente estimado es grande: un aumento de un punto (sobre 10) en la confianza en el sistema político correlaciona con un aumento de 0,29 puntos (sobre 10) en la preferencia por subir los impuestos. Esta variable es estadísticamente significativa y aumenta el poder explicativo de 0,16 a 0,24. La correlación de la variable dependiente, con el resto de variables consideradas (ingresos, sociodemográficas, políticas) se mantiene, en líneas generales, similar tras controlar por la confianza.

El gráfico 4 representa gráficamente estos resultados, mostrando los coeficientes de la regresión de la preferencia por subir impuestos sobre las distintas variables analizadas. Con el fin de poder comparar las magnitudes, todas las variables explicativas se han estandarizado, de tal manera que los coeficientes indican el efecto de aumentar las correspondientes variables independientes en una desviación estándar. El gráfico pone de relieve que la confianza en el sistema político correlaciona fuertemente con la preferencia por subir impuestos, seguida de las variables de intención de voto.

Los cuadros 3 y 4 muestran los resultados de regresiones análogas a las comentadas hasta ahora, pero en las que la variable dependiente es dar una opinión positiva (cuadro n.º 3) o negativa (cuadro n.º 4) sobre los impuestos. En líneas genera-

CUADRO N.º 2

PREFERENCIAS POR SUBIR IMPUESTOS

	(1) SUBIR IMPUESTOS	(2) SUBIR IMPUESTOS	(3) SUBIR IMPUESTOS	(4) SUBIR IMPUESTOS
Quintil 2	0,24 (0,16)	0,17 (0,16)	0,11 (0,15)	0,18 (0,15)
Quintil 3	0,39*** (0,15)	0,29* (0,15)	0,27* (0,15)	0,44*** (0,14)
Quintil 4	0,72*** (0,16)	0,57*** (0,17)	0,52*** (0,16)	0,62*** (0,15)
Quintil 5	0,96*** (0,17)	0,77*** (0,18)	0,63*** (0,16)	0,75*** (0,16)
Mujer		-0,26** (0,11)	-0,36*** (0,10)	-0,45*** (0,09)
Edad [31,65]		0,05 (0,13)	0,11 (0,12)	0,16 (0,12)
Edad >65		0,57** (0,22)	0,75*** (0,20)	0,72*** (0,18)
Secundaria		0,37** (0,16)	0,24* (0,15)	0,12 (0,14)
Bachillerato		0,34* (0,18)	0,32* (0,17)	0,14 (0,16)
FP		0,31 (0,20)	0,30 (0,19)	0,13 (0,18)
Superior		0,46*** (0,15)	0,54*** (0,14)	0,40*** (0,13)
Voto UP			0,63*** (0,19)	0,70*** (0,19)
Voto Cs			-1,24*** (0,20)	-0,51*** (0,18)
Voto PP			-2,17*** (0,18)	-1,21*** (0,18)
Voto VOX			-2,79*** (0,18)	-1,63*** (0,19)
Voto Otro			-1,50*** (0,13)	-0,69*** (0,13)
Confianza				0,29*** (0,02)
N	4.566	4.566	4.566	4.566
R2	0,02	0,02	0,16	0,24

Notas: Cada columna muestra los resultados de una regresión de mínimos cuadrados ordinarios de la variable *Subir Impuestos* sobre las variables indicadas en las filas. * $p < 0,10$. ** $p < 0,05$. *** $p < 0,01$. Errores estándar robustos entre paréntesis. Pesos muestrales utilizados.

IV. CONCLUSIONES

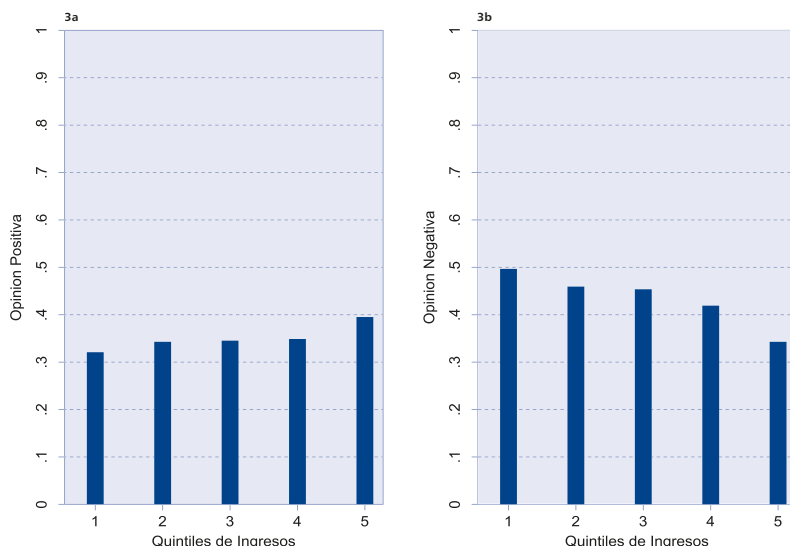
En este artículo han sido analizadas las preferencias de los españoles sobre el nivel impositivo a través de una encuesta *online* llevada a cabo tras el comienzo de la pandemia de COVID-19. De manera quizá un tanto sorprendente, los individuos de mayor renta expresan una mayor preferencia por aumentar impuestos, incluso después de controlar por una serie de factores sociodemográficos y políticos. Asimismo, los hombres y las personas de más edad y más nivel educativo, y los votantes de izquierdas expresan una mayor preferencia por subir los impuestos. Por último, una mayor confianza en el sistema político está fuertemente asociada a una mayor preferencia por subir los impuestos y a mencionar opiniones positivas sobre los mismos aun después de controlar por el resto de factores económicos, sociodemográficos y políticos.

Resulta interesante comparar estos resultados con los de Gethin, Martínez-Toledano y Piketty (2021). Estos autores muestran, usando datos electorales de 21 democracias occidentales, que la educación está positivamente correlacionada con votar a partidos de izquierdas (a diferencia de lo que ocurría en los años cincuenta o sesenta del pasado siglo) mientras que la renta correlaciona con votar más a partidos de derechas, al igual que en el pasado. Sin embargo, los resultados son algo distintos para España, donde los votantes más educados (así como los de más renta) siguen votando más a la derecha (si bien esta asociación es menor que en décadas anteriores). Sin embargo, los resultados del presente artículo indican que la renta y la educación

les, los resultados son similares a los obtenidos para la variable *Subir Impuestos*. Los individuos que viven en hogares de mayor renta son más (menos) propensos a dar una opinión positiva (negativa) sobre los impuestos, aunque las correlaciones son algo más tenues que para *Subir Impuestos*. Los hombres, los in-

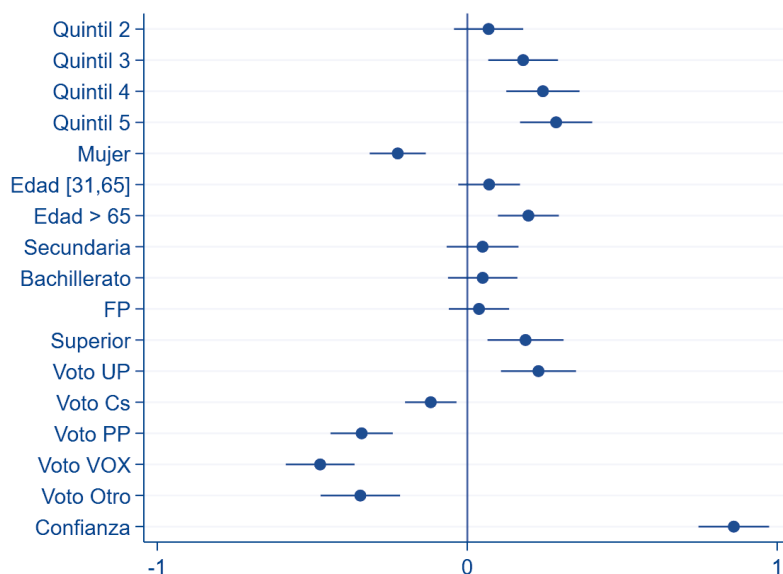
dividuos de más edad y de más nivel educativo y los votantes del PSOE tienen una opinión más favorable de los impuestos. Por último, la confianza en el sistema político está fuertemente correlacionada con tener una opinión positiva de los impuestos aun después de controlar por el resto de factores.

GRÁFICO 3
OPINIONES SOBRE LOS IMPUESTOS (RESUMEN), POR INGRESOS



Notas: Las barras muestran la proporción de encuestados que, cuando expresan su opinión sobre los impuestos, dan al menos una respuesta positiva y ninguna negativa (gráfico 3a) o al menos una respuesta negativa y ninguna positiva (gráfico 3b). Los encuestados se agrupan en cinco quintiles de acuerdo con sus ingresos familiares per cápita en 2019. El quintil 1 es el quintil más pobre y el 5, el más rico.
Fuente: Elaboración propia.

GRÁFICO 4
PREFERENCIA POR SUBIR LOS IMPUESTOS: COEFICIENTES



Notas: Los puntos muestran los coeficientes de la regresión de la preferencia media, del 0 al 10, sobre subir (10) o bajar (0) los impuestos sobre las variables indicadas (cuadro 2, columna 4), donde las variables explicativas han sido estandarizadas (media=0, desviación estándar=1). Las líneas denotan intervalos de confianza al 95 por 100. Quintil 2-5 denota quintiles de renta familiar, donde 5 es el más rico. Confianza denota confianza en el sistema político.
Fuente: Elaboración propia.

están positivamente correlacionadas con una mayor preferencia por subir los impuestos, postura que normalmente se asocia a la izquierda. Hay varias posibles explicaciones a esta aparente discrepancia. Una es que, obviamente, hay muchos más factores además de los impuestos que condicionan el voto. Otra es el marco temporal: los resultados de Gethin, Martínez-Toledano y Piketty (2021) cubren las elecciones de la década de 2010 mientras que los del presente artículo se basan en una encuesta realizada en 2020. Dado el cambio de tendencia en la asociación entre voto y educación mostrada por Gethin, Martínez-Toledano y Piketty (2021), es posible que los resultados del presente artículo anticipen un cambio de signo de esa asociación, en línea con lo ocurrido en otros países.

Este artículo sugiere varias líneas de investigación. En primer lugar, es importante tratar de hallar cuál es el mecanismo que genera la correlación positiva entre renta y preferencia por subir los impuestos. Una posibilidad es que los impuestos no sean percibidos por los individuos (tanto ricos como pobres) como una herramienta redistributiva; por ejemplo, si los votantes pobres piensan que los aumentos impositivos les supondrán un mayor coste fiscal directo, pero no se beneficiarán de gastos o transferencias como resultado de esos aumentos, dichos votantes probablemente estarán en contra de aumentar los impuestos. En segundo lugar, la encuesta se hizo en el contexto específico de la pandemia por lo que sería importante examinar la evolución temporal de las correlaciones observadas y si estas pueden verse afectadas por determinados shocks. No obstante,

CUADRO N.º 3

PREFERENCIAS POR SUBIR IMPUESTOS

	(1) OPINIÓN POSITIVA	(2) OPINIÓN POSITIVA	(3) OPINIÓN POSITIVA	(4) OPINIÓN POSITIVA
Quintil 2	0,02 (0,03)	0,01 (0,03)	-0,00 (0,03)	0,01 (0,03)
Quintil 3	0,02 (0,03)	-0,00 (0,03)	-0,00 (0,03)	0,02 (0,03)
Quintil 4	0,03 (0,03)	-0,01 (0,03)	-0,02 (0,03)	-0,01 (0,03)
Quintil 5	0,07** (0,03)	0,02 (0,03)	0,01 (0,03)	0,02 (0,03)
Mujer		-0,02 (0,02)	-0,03 (0,02)	-0,04** (0,02)
Edad [31,65]		0,01 (0,02)	0,02 (0,02)	0,02 (0,02)
Edad >65		0,11*** (0,04)	0,11*** (0,04)	0,11*** (0,04)
Secundaria		0,07*** (0,03)	0,06** (0,03)	0,04 (0,03)
Bachillerato		0,07** (0,03)	0,07** (0,03)	0,04 (0,03)
FP		0,07* (0,04)	0,06* (0,04)	0,04 (0,04)
Superior		0,14*** (0,03)	0,14*** (0,03)	0,12*** (0,03)
Voto UP			-0,07** (0,03)	-0,06* (0,03)
Voto Cs			-0,22*** (0,04)	-0,13*** (0,04)
Voto PP			-0,27*** (0,04)	-0,15*** (0,04)
Voto VOX			-0,39*** (0,03)	-0,25*** (0,03)
Voto otro			-0,27*** (0,02)	-0,17*** (0,03)
Confianza				0,04*** (0,00)
N	4.566	4.566	4.566	4.566
R2	0,00	0,02	0,09	0,13

Notas: Cada columna muestra los resultados de una regresión de mínimos cuadrados ordinarios de la variable *Opinión positiva* sobre las variables indicadas en las filas. * $p < 0,10$. ** $p < 0,05$. *** $p < 0,01$. Errores estándar robustos entre paréntesis. Pesos muestrales utilizados.

das, no reveladas a partir de decisiones reales de los individuos. Sería importante que trabajos futuros miraran cómo estas preferencias operan en la vida real. Una manera de aproximarse a esto sería mediante el estudio de experimentos de laboratorio en los que los pagos a los individuos dependen de sus decisiones.

NOTAS

(*) Los análisis, opiniones y conclusiones de estas investigaciones representan las ideas del autor, con las que no necesariamente coincide el Banco de España. Agradezco a LAURA HOSPIDO, FRANCISCO PEDRAJA-CHAPARRO y PILAR SORRIBAS-NAVARRO sus muy útiles comentarios.

(1) Para más detalles sobre la encuesta, consultar MARTINEZ-BRAVO y SANZ (2021) y MARTINEZ-BRAVO y SANZ (2021b). El primer artículo utiliza la encuesta para analizar los cambios en la distribución de la renta y el bienestar emocional tras la pandemia. El segundo artículo estima, mediante un experimento, el efecto de la gestión de la pandemia sobre la confianza en las instituciones y la voluntad de cumplir con las recomendaciones sanitarias (por ejemplo, vacunarse).

(2) El marco teórico de MELTZER y RICHARD (1981) tiene una única variable, el tipo impositivo, que captura tanto el nivel impositivo como la redistribución. Es decir, este marco no permite separar ambos conceptos.

(3) También hay estudios que muestran que el tamaño del sector público y la redistribución dependen no solo de las preferencias de los ciudadanos, sino también del sistema de gobierno o electoral utilizado. En este sentido, IVERSEN y SOSKICE (2006) muestran que los sistemas de representación proporcional conducen a más redistribución que los mayoritarios. Para el caso de España, SANZ (2020) muestra que un sistema de democracia directa reduce los gastos e ingresos públicos, comparado con un sistema de democracia representativa.

(4) Un análisis de las causas de la (falta de) confianza va más allá del alcance de este artículo. Para el caso de España, hay evidencia de que la corrupción juega un papel importante en el deterioro de la confianza: ver la evidencia de SORRIBAS-NAVARRO (2021) así como SANZ, SOLÉ-OLLÉ y SORRIBAS-NAVARRO (2021), que muestran fuertes impactos de la corrupción en el voto.

(5) Las áreas Nielsen son seis zonas geográficas, de características relativamente homogéneas, en las que YouGov divide el territorio español.

es reseñable que no se observan cambios entre las olas de mayo y noviembre de 2020 (mientras que sí se observan grandes diferencias en ingresos y bienestar emocional; ver Martínez-Bravo y Sanz, 2021). En tercer lugar, el presente artículo se centra en las preferencias impositivas. Sería

importante estudiar también conjuntamente el lado del gasto público y cómo las preferencias sobre impuestos y gasto interactúan. Por último, el presente artículo tiene las limitaciones inherentes a las encuestas; quizá de forma más notable, las preferencias son tendencias expresa-

CUADRO N.º 4

OPINIÓN NEGATIVA SOBRE LOS IMPUESTOS

	(1) OPINIÓN NEGATIVA	(2) OPINIÓN NEGATIVA	(3) OPINIÓN NEGATIVA	(4) OPINIÓN NEGATIVA
Quintil 2	-0,04 (0,03)	-0,02 (0,03)	-0,01 (0,03)	-0,02 (0,03)
Quintil 3	-0,04 (0,03)	-0,01 (0,03)	-0,01 (0,03)	-0,03 (0,03)
Quintil 4	-0,08*** (0,03)	-0,03 (0,03)	-0,02 (0,03)	-0,03 (0,03)
Quintil 5	-0,15*** (0,03)	-0,09*** (0,03)	-0,07** (0,03)	-0,08*** (0,03)
Mujer		0,08*** (0,02)	0,09*** (0,02)	0,10*** (0,02)
Edad [31,65]		0,02 (0,02)	0,01 (0,02)	0,01 (0,02)
Edad >65		-0,09** (0,04)	-0,09** (0,04)	-0,08** (0,04)
Secundaria		-0,06** (0,03)	-0,04 (0,03)	-0,02 (0,03)
Bachillerato		-0,10*** (0,03)	-0,09*** (0,03)	-0,07** (0,03)
FP		-0,07* (0,04)	-0,07* (0,04)	-0,05 (0,04)
Superior		-0,15*** (0,03)	-0,15*** (0,03)	-0,13*** (0,03)
Voto UP			-0,04 (0,03)	-0,05 (0,03)
Voto Cs			0,20*** (0,04)	0,11** (0,04)
Voto PP			0,23*** (0,03)	0,11*** (0,04)
Voto VOX			0,36*** (0,03)	0,22*** (0,04)
Voto otro			0,28*** (0,02)	0,18*** (0,02)
Confianza				-0,04*** (0,00)
N	4.566	4.566	4.566	4.566
R2	0,01	0,03	0,11	0,14

Notas: Cada columna muestra los resultados de una regresión de mínimos cuadrados ordinarios de la variable *Opinión negativa* sobre las variables indicadas en las filas. * $p < 0,10$. ** $p < 0,05$. *** $p < 0,01$. Errores estándar robustos entre paréntesis. Pesos muestrales utilizados.

(6) RÍOS FERNÁNDEZ y VALLE MORENO (2021) encuentran, utilizando encuestas del CIS de 2015 a 2020, un progresivo aumento del porcentaje de población que piensa que se pagan pocos impuestos, y una progresiva caída de los que opinan que se pagan muchos impuestos.

(7) Esto no es sorprendente dado que la muestra está diseñada para replicar la distribución de la población adulta española por edad, género, nivel educativo y área geográfica. Por tanto, los pesos únicamente corrigen

pequeñas desviaciones a la hora de llenar las cuotas en el muestreo.

(8) El texto completo de la pregunta es: «La crisis de la COVID-19 ha generado un aumento en el gasto público para financiar las medidas sanitarias y de protección social, como los ERTE. Algunos políticos argumentan que es necesario subir los impuestos para financiar estos costes adicionales, mientras que otros proponen bajar los impuestos para reactivar la economía. Indica tu postura en

una escala del 0 al 10, donde 0 es “bajar impuestos” y 10 “subir impuestos”».

(9) Más en concreto, se pregunta a los encuestados por sus ingresos familiares netos totales, es decir, incluyendo salarios, rentas de actividades profesionales, pensiones y transferencias públicas, durante el mes medio de 2019. Los encuestados tenían que elegir el intervalo en el que caían sus ingresos familiares: 0-300, 300-600, 601-900, 901-1.200, 1.201-1.500, 1.501-1.800, 1.801-2.100, 2.101-2.300, 2.401-3.000, 3.001-4.500, 4.501-6.000, más de 6.000. Tomo el punto medio de cada intervalo como *proxy* del nivel de ingresos, y 8.000 para el superior. Para comparar entre hogares, sigo la convención de Eurostat y asigno un peso de 0,5 a los niños. La variable *Ingresos* se obtiene de dividir los ingresos totales reportados por el encuestado por el número de individuos en el hogar (multiplicando por 0,5 a los niños) y multiplicando por tres, lo que corresponde a un hogar de dos adultos y dos niños.

(10) La variable *Voto a Otros* no tiene una interpretación clara ya que incluye una gran variedad de partidos distintos (y la abstención).

BIBLIOGRAFÍA

- ACEMOGLU, D., CHEEMA, A., KHWAJA, A. I. y ROBINSON, J. A. (2020). Trust in state and nonstate actors: Evidence from dispute resolution in Pakistan. *Journal of Political Economy*, 128(8), pp. 3090-3147.
- ALESINA, A. y GIULIANO, P. (2011). Preferences for redistribution. En J. BENHABIB y A. BISIN (eds.), *Handbook of Social Economics*, pp. 93-131. Amsterdam: North Holland.
- ALESINA, A. y GLAESER, E. L. (2004). *Fighting Poverty in the US and Europe: A World of Difference*. Oxford: Oxford University Press.
- ALESINA, A. y LA FERRARA, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6), pp. 897-931.
- ALGAN, Y. y CAHUC, P. (2010). Inherited trust and growth. *American Economic Review*, 100(5), pp. 2060-2092.
- BENABOU, R. y TIROLE, J. (2006). Belief in a just world and redistributive politics. *The Quarterly Journal of Economics*, 121(2), pp. 699-746.
- CABRERA-ÁLVAREZ, P. y ESCOBAR, M. (2019). El efecto de la ponderación y

la imputación en el sesgo de los estudios electorales en España. <i>Revista Española de Investigaciones Sociológicas</i>, 165, pp. 45-64. FONG, CH. (2001). Social preferences, self-interest, and the demand for redistribution. <i>Journal of Public Economics</i>, 82(2), pp. 225-246. GETHIN, A., MARTÍNEZ-TOLEDANO, C. y PIKETTY, TH. (2021). Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020. <i>The Quarterly Journal of Economics</i>; qjab036, https://doi.org/10.1093/qje/qjab036 IVERSEN, T. y SOSKICE, D. (2006). Electoral institutions and the politics of coalitions: Why some democracies redistribute more than others. <i>American Political Science Review</i>, 100(2), pp. 165-181. KENWORTHY, L. y MCCALL, L. (2008). Inequality, Public Opinion and	Redistribution. <i>Socio-Economic Review</i>, 6(1), pp. 35-68. KUZIEMKO, I., NORTON, M. I., SÁEZ, E. y STANTCHEVA, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. <i>American Economic Review</i>, 105(4), pp. 1478-1508. MARTINEZ-BRAVO, M. y SANZ, C. (2021a). Inequality and Psychological Well-being in Times of COVID-19: Evidence from Spain. <i>SERIES-Journal of the Spanish Economic Association</i>, Volume 12, December 2021, pp. 489-548. MARTINEZ-BRAVO, M. y SANZ, C. (2021b). The Management of the Pandemic and its Effects on Trust and Accountability. Unpublished Manuscript. MELTZER, A. y RICHARD, S. (1981). A rational theory of the size of government. <i>Journal of Political Economy</i>, 89, pp. 914-927.	PIKETTY, T. (1995). Social mobility and redistributive politics. <i>The Quarterly Journal of Economics</i>, 110(3), pp. 551-584. RÍOS FERNÁNDEZ, M. y VALLE MORENO, F. (2021). COVID-19, intervención pública y preferencias redistributivas: un análisis descriptivo del caso español. <i>Clivatge</i> (9), e-34368. https://doi.org/10.1344/CLIVATGE2021.9.10 SANZ, C. (2020). Direct democracy and government size: Evidence from Spain. <i>Political Science Research and Methods</i>, 8(4), pp. 630-645. SANZ, C., SOLÉ-OLLÉ, A. y SORRIBAS-NAVARRO, P. (2021). Betrayed by the elites: how corruption amplifies the political effects of recessions. <i>Comparative Political Studies</i>. Forthcoming. SORRIBAS-NAVARRO, P. (2021). Corrupción y confianza en los políticos en España. <i>Papeles de Economía Española</i>, (168), pp. 157-191.
--	---	---

EVALUACIÓN DEL IMPACTO DE UN INCREMENTO DEL IVA REDUCIDO SOBRE EL GASTO DE LOS HOGARES EN ESPAÑA

Ignacio MORAL-ARCE

Instituto de Estudios Fiscales

Resumen

Este trabajo analiza el impacto que un incremento del tipo de gravamen en el impuesto sobre el valor añadido de tipo reducido en los años 2010 y 2012 de 1 y 2 puntos, respectivamente, ha tenido sobre el gasto de los hogares en España. Para ello se utiliza el método de regresión en discontinuidad y se realizan cálculos para diferentes tipos de bienes y servicios y tipos de hogares en función del nivel de ingreso. Los resultados muestran que la reforma del año 2012 tiene efectos significativos sobre el gasto de los hogares, con reducciones generales de gasto, y la de 2010 incide más en familias con bajos ingresos.

Palabras clave: evaluación de políticas públicas, evaluación de impacto, regresión en discontinuidad, impuestos indirectos, IVA.

Abstract

This paper analyzes the impact that an increase in the tax rate on the Value Added Tax (VAT) in 2010 and 2012 by 1 and 2 points respectively has had on household expenditure in Spain. We use the discontinuity regression method and we show calculations for different types of goods and services and levels of household income. The results reveal that the 2012 reform has significant effects on households, with general expenditure reductions, and the 2010 reform has more impact on low-income families.

Keywords: public policy evaluation, impact evaluation, discontinuity regression, indirect taxes, VAT.

JEL classification: C21, D12, H31.

I. INTRODUCCIÓN

La generación de evidencia para la toma de decisiones políticas está adquiriendo una gran importancia en los últimos años, no solo desde el ámbito de la gestión pública, sino también desde el mundo académico y la propia sociedad mediante la rendición de cuentas. La evaluación de políticas públicas permite generar hallazgos y resultados que el gestor público utiliza, y que le permite reducir, en cierto grado, la gran incertidumbre existente a la hora de la toma de decisiones, logrando de este modo proponer medidas más eficientes y eficaces.

Tanto desde el lado del gasto como del ingreso, las alternativas que existen para ofrecer información al diseñador de una intervención pública son múltiples, factibles y compatibles, pudiendo ir desde ópticas más cualitativas hasta cuantitativas, pasando por enfoques mixtos, y abarcando todo el ciclo de una política pública: diseño, ejecución y finalización.

Por lo que se refiere a la última fase, las evaluaciones se centran, fundamentalmente, en detectar el resultado que ha tenido una intervención pública en una variable de interés, que determina el grado de éxito de este programa, siendo la evaluación del impacto el método más habitual.

A la hora de determinar el efecto de una determinada política se pueden considerar dos grandes aproximaciones: un enfoque macroeconómico, que permite estimar el efecto sobre variables agregadas de la economía, como es el PIB, y que emplean modelos de simulación con diferente grado de complejidad y supuestos, abarcando desde opciones más parciales como el análisis de series temporales, hasta aproximaciones *input/output* o modelos de equilibrio general. La otra alternativa es un enfoque microeconómico, donde las posibilidades de evaluación son dos. Por un lado, modelos de microsimulación a partir de diferentes tipos de ecuaciones econométricas, que evalúan el efecto de reformas políticas en diferentes individuos, familias, empresas, y ofreciendo resultados, no solo del impacto promedio, sino por tipología de agentes y analizando la progresividad, equidad y redistribución entre otros conceptos. La otra opción para realizar una evaluación del impacto con un enfoque micro es emplear contrafactuales, utilizando grupos de tratamiento y control.

En este contexto, el objetivo de este trabajo es evaluar de manera conjunta el efecto de un incremento del tipo reducido del IVA de 1 y 2 puntos realizados en julio de 2010 y septiembre de 2012, estudiando en qué medida dicho incremento afectó al gasto de bienes y servicios que realizan los ho-

gares, mediante la utilización de métodos cuasiexperimentales. Existen estudios que han analizado los efectos de la reforma del IVA de 2012 (1). A diferencia de la mayoría de trabajos previos y debido a la disponibilidad de datos mensuales sobre el gasto en bienes y servicios de los hogares, en esta investigación se aplica la técnica de evaluación de regresión en discontinuidad para determinar el efecto del incremento de gravamen (2). Además, se realizan estimaciones del efecto que supone un incremento impositivo sobre el gasto, tanto en el margen extensivo como intensivo, valorando finalmente cómo el aumento del IVA impacta sobre determinada tipología de hogares o bienes.

Las novedades que incorpora esta investigación respecto a la literatura existente son fundamentalmente tres: 1) es la primera que analiza simultáneamente las dos subidas de los tipos del IVA reducido en el período de poco más de dos años, aplicando una variante del método de regresión en discontinuidad diseñado especialmente para estas situaciones; 2) evalúa el impacto de la reforma impositiva sobre el consumo de los hogares tanto en el margen extensivo como en el intensivo; y 3) estudia, desde un punto de vista de análisis de heterogeneidad, el impacto que en el gasto de di-

ferentes tipos de hogares, así como distintos bienes y servicios, ha tenido la reforma.

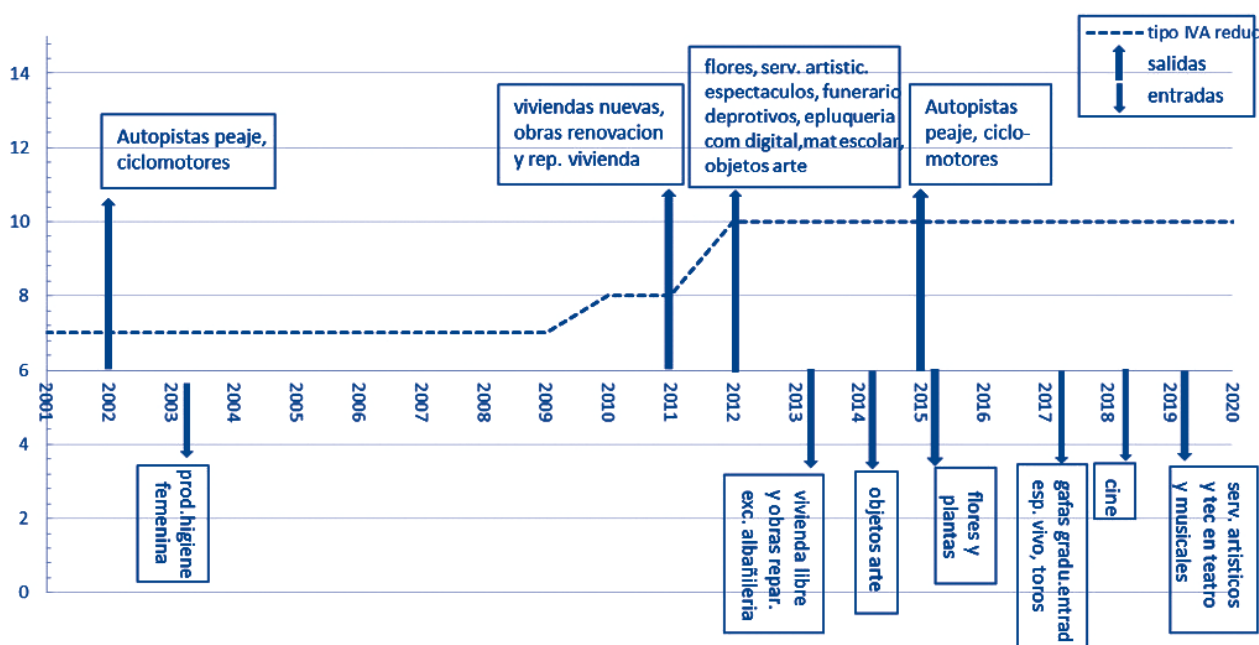
La estructura del trabajo es la siguiente. En la siguiente sección se realiza un resumen sobre las principales modificaciones del IVA que han existido en España en los últimos años. A continuación, se repasa la literatura empírica que analiza el efecto de diferentes reformas de impuestos indirectos en España, como el IVA y los impuestos especiales (II. EE.). En la cuarta sección, se cuantifica si el incremento de gravamen en el IVA reducido supuso un cambio en el gasto de los hogares, utilizando el método de regresión en discontinuidad, concluyendo la sección con un análisis de heterogeneidad. Por último, se dedica una sección a destacar las principales conclusiones obtenidas.

II. IVA REDUCIDO Y CAMBIOS NORMATIVOS

En el año 1986 España introdujo el IVA con diferentes tipos impositivos (general, reducido y superreducido), lo que permitía influir sobre la distribución de la renta de los hogares y facilitar el acceso de la población a determinados productos. La regla de imposición óptima de Ramsey determina

GRÁFICO 1

PRINCIPALES CAMBIOS EN EL IVA REDUCIDO: ENTRADAS Y SALIDAS DE PRODUCTOS Y CAMBIOS DE TIPOS



que deben gravarse con tipos más altos los bienes cuya demanda es más inelástica. Sin embargo, los problemas de diseño de un impuesto que grave el consumo de bienes y servicios en función de sus respectivas elasticidades-precio son considerables, ya que la Administración Pública no solamente debe ser capaz de estimar las elasticidades de cada producto, sino que debe realizar este cálculo con regularidad debido a los cambios en las preferencias de los individuos y a las variaciones de las elasticidades-precio cruzadas como consecuencia de la aparición de nuevos bienes en el mercado.

A lo largo de las últimas décadas el IVA ha sufrido diferentes modificaciones, tanto el general como el reducido y superreducido, pero especialmente el IVA reducido es el que mayor número de cambios ha experimentado, desde la entrada y salida de varios productos y servicios, así como variación de tipos. Por lo que se refiere a este último, para tratar de contrarrestar la negativa evolución que estaban teniendo las finanzas públicas españolas a partir de la crisis económica de 2008, además de ciertas medidas destinadas a la contención del gasto, se estableció un aumento en la mayoría de los impuestos (Sanz y Romero, 2012a). Por lo que se refiere al IVA, los principales cambios normativos que se produjeron en los años 2010 y 2012 fueron el incremento de tipos en el IVA reducido y general, con aumento de tipos en el primero del 7 al 8 por 100 en 2010, y al 10 por 100 en 2012, mientras que el tipo general experimentó sendas subidas de 2 y 3 puntos porcentuales en esas fechas, respectivamente, pasando del 16 al 21 por 100.

Por lo que respecta al IVA reducido, objeto de análisis en este trabajo, en el gráfico 1 se muestra, de forma simultánea, las grandes modificaciones en este tipo del IVA, tanto de salidas como entradas de productos y servicios afectados, así como la variación del gravamen existente.

Se observa que los grandes cambios se sitúan en la última década, a partir de 2010, tanto con subidas de tipos como por la inclusión y exclusión de diferentes bienes y servicios.

III. EVALUACIONES DE LAS REFORMAS DEL IVA EN ESPAÑA

A partir de los noventa, existen estudios que examinan el efecto de las reformas impositivas de impuestos indirectos, IVA e II. EE., analizando

tanto efectos recaudatorios como redistributivos o impactos en la desigualdad. La mayoría de estos trabajos que evalúan el efecto de los impuestos indirectos se pueden agrupar en dos categorías: a) los que emplean métodos de simulación, ya sea con una óptica macro mediante el uso de modelos de equilibrio general, o modelos micro, principalmente mediante la estimación de modelos de demanda, y el cálculo de diferente tipo de elasticidades; y b) evaluaciones del impacto que emplean grupo de control con diseños cuasiexperimentales.

Por lo que se refiere a la estimación y simulación de reformas de impuestos indirectos, con una óptica macro, destacan los trabajos de series temporales de Izquierdo, Melguizo y Taguas (2001) con modelos ARIMA que estiman el efecto que los cambios normativos en el IVA y en los II. EE. de alcohol, tabaco e hidrocarburos han tenido en los precios, empleando datos mensuales para el período 1983 a 2000, similar a Bello y Contín-Pilart (2012) con series temporales mensuales. También Martínez-Vázquez y Bird (2010) analizan el efecto del IVA empleando un panel de datos, con información de países y empleando el método generalizado de momentos. Otra alternativa son los modelos de equilibrio general (MEG) donde destacan el trabajo de Kehoe (1989) que estudia el efecto de introducir el IVA en España en 1986 sobre los precios, actividad y desempleo entre otras variables, o los de Fernández de Córdoba y Torres (2010) y Conesa *et al.* (2010) que evalúan con un MEG dinámico el efecto del aumento del IVA en 2010 sobre diferentes variables macro (3).

Por lo que se refiere a la aproximación micro, la mayoría de los trabajos presentan ciertas características similares. Primero, la fuente de información que se emplea es la *Encuesta de presupuestos familiares (EPF)* del INE, que permite realizar análisis a nivel de hogares. Segundo, la variable de interés analizada es el gasto total de los hogares como medida de capacidad económica de estos (4). Por último, para la estimación se emplean modelos de demanda, más o menos complejos, desde ecuaciones unidimensionales de demanda para un determinado producto o servicio hasta sistema de ecuaciones simultáneas como los sistemas de modelos ideales de demanda (*almost ideal demand system*, *AIDS*) (5). Considerando la amplia literatura en esta área, se destacan, entre otros, los trabajos de Labeaga y López (1994, 1996) que estiman el efecto de la reforma del IVA de los años 1992 y 1995 respectivamente; Romero (2014), que evalúa diferentes tipos de reformas; mientras que

Serrano y Oliva (2011) estudian los efectos redistributivos de la subida de tipos del IVA en 2010. Para los bienes culturales, García-Enríquez y Echevarría (2018) estiman el efecto del incremento del IVA en 2012. Finalmente, en II.EE. cabe mencionar los trabajos de Asensio, Matas y Raymond (2003) con datos de la EPF de 1990 evaluando el efecto redistributivo de los impuestos a los combustibles, o López, González y Fellingner (2003) que examina el impacto del aumento de los impuestos especiales (6).

La última de las opciones a la hora de estimar el efecto que tienen las reformas del IVA son las evaluaciones del impacto empleando contrafactuales. Ateca-Amestoy, Gardeazabal y Ugidos (2020), con datos de la EPF, realizan una evaluación de impacto, empleando el método de diferencias en diferencias (*dif-in-dif*), del efecto de un aumento del IVA del cine y las artes escénicas en 2012 en la Península, utilizando como grupo de control las Islas Canarias que no forman parte del territorio del IVA español y donde la tributación indirecta, que grava el cine y las artes escénicas a través del impuesto general indirecto de Canarias (IGIC), había variado en julio de 2012 en solo 2 puntos porcentuales. De manera complementaria, Moral-Arce y Gómez-Antonio (2020) estudian el efecto de una reducción del tipo de gravamen del IVA en bienes y servicios culturales empleando el método de *dif-in-dif* y regresión en discontinuidad. El trabajo de la AIREF (2020) analiza el efecto que la subida de tipos en el IVA en 2012 tiene en el gasto de los hogares empleando el método de regresión en discontinuidad, mientras que Artés, Botello y Sánchez-Fuentes (2019) estudian el impacto que las reformas del IVA de 2010 y 2012 tienen en la disposición de los consumidores a comprar utilizando las búsquedas de Google como *proxy* y empleando el método de regresión en discontinuidad, estimando efectos en las búsquedas sobre bienes duraderos. Bajo-Buenestado (2017) estima el impacto de una reducción del impuesto a la gasolina en los precios en el surtidor con datos de precios diarios de estaciones de servicio y empleando un *dif-in-dif*, obteniendo distintos grados de transferencia a los precios minoristas (7).

IV. ESTIMACIÓN DEL IMPACTO DEL INCREMENTO DE TIPOS DEL IVA REDUCIDO EN EL GASTO DE LOS HOGARES

En esta sección se evalúan las reformas de julio de 2010 y septiembre de 2012 que supusieron un

incremento del tipo impositivo del IVA reducido de 1 y 2 puntos respectivamente en el gasto de los hogares en bienes y servicios afectados por este tipo del IVA. El objetivo es analizar si afectó al número de hogares que consumían este tipo de bienes (margen extensivo) y al gasto medio del hogar (margen intensivo) (8).

1. Efecto sobre el número de hogares que consumen bienes y servicios (margen extensivo)

El objetivo en esta sección es determinar si la subida de gravamen del IVA reducido en 2010 y 2012 influyó sobre el número de hogares que consumían los bienes y servicios afectados por este tipo del IVA. Para el análisis, la fuente de información que se empleará para estimar una regresión en discontinuidad procede de la *Encuesta de presupuestos familiares (EPF)* para el período 2007-2015 y de un fichero de enlace facilitado por el INE, que recoge el mes en que se produjo la recopilación de información que aportaban los hogares a la EPF y que permite obtener series de gasto en bienes y servicios mensuales. Para el período analizado, el INE también provee el IPC mensual por tipo de bien (codificación COICOP) y por comunidad autónoma con un mayor grado de desagregación. A partir de la información a nivel de hogar, se construye la variable dicotómica objeto de estudio $I(\text{Gasto}_{hst} > 0)$ que indica si el hogar «*h*» en el mes «*t*» ha realizado gastos en bienes y servicios «*s*» afectados por el IVA reducido:

$$Phogbbss_{st} = \frac{n^{\circ} \text{hogares consumen bbss grupo "s" en el mes "t"}}{n^{\circ} \text{total de hogares en momento "t"}} = \frac{\sum_{i=1}^{n_t} I(\text{Gasto}_{hst} > 0)}{n_{ht}} \quad [1]$$

donde $I(\text{Gasto}_{hst} > 0)$ es una variable dicotómica que toma el valor «1» si el hogar «*h*» en el mes «*t*» realiza algún gasto en bienes y servicios afectados por el IVA reducido, que han sido agrupados en (*s*=alimentos, bebidas, restaurantes, hostelería, transporte) y n_{ht} es el total de hogares en el mes «*t*». Los gráficos 2 a 4 muestran el porcentaje de hogares que realizan algún gasto de este tipo de bienes y servicios, denominado $Phogbbss_{st}$.

Los diferentes gráficos muestran el estimador no paramétrico, empleando polinomios locales de tercer orden, de la proporción de los hogares que gastan en los bienes afectados por el IVA reducido calculado en cada uno de los tres tramos relevantes: antes de julio 2010, entre julio de 2010 hasta septiembre de 2012, y posterior a esa fecha.

GRÁFICO 2

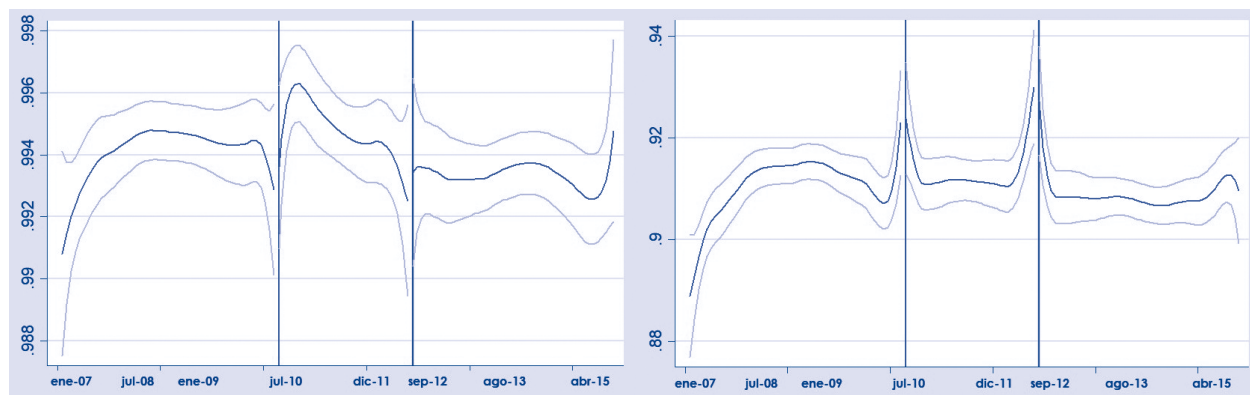
PROMEDIO DE HOGARES CON GASTO EN ALIMENTOS (IZQD.) Y BEBIDAS NO ALCOHOL (DCHA.)

GRÁFICO 3

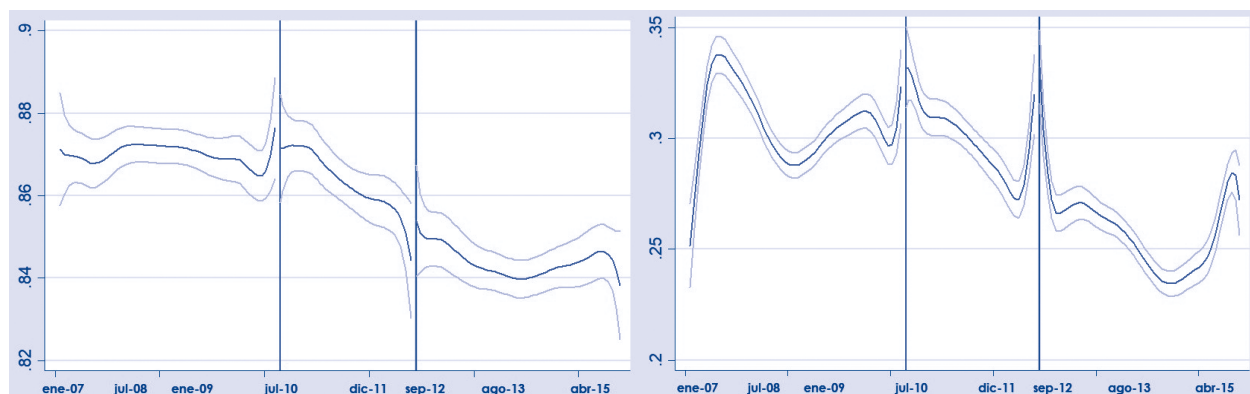
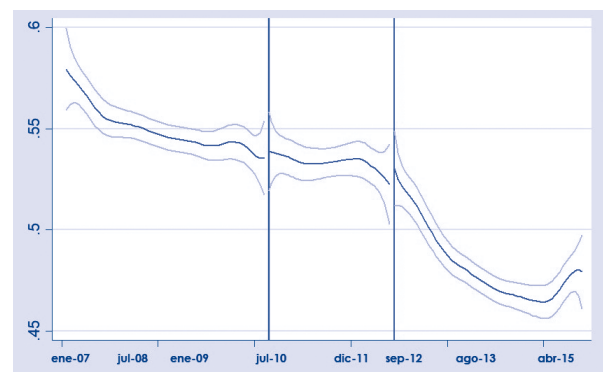
PROMEDIO DE HOGARES CON GASTO EN RESTAURACIÓN (IZQD.) Y HOSTELERÍA (DCHA.)

GRÁFICO 4

PROMEDIO DE HOGARES CON GASTO EN TRANSPORTE

Durante el período 2007-2015 casi la totalidad de los hogares consumieron alimentos afectados por tipos reducidos, con valores por encima del 99 por 100, y alrededor del 90 por 100 para bebidas no alcohólicas, con valores que permanecen bastante estables durante estos años, mientras que restauración se sitúa en torno a 85 por 100 con un ligero descenso a lo largo del período, observándose valores muchos más bajos en hostelería y transporte, con promedios para el período estudiado del 28 por 100 y 50 por 100, respectivamente, y con un marcado descenso en ambas series a lo largo del tiempo. Los gráficos también muestran que el porcentaje de hogares que consumen estos bienes después de julio de 2010 o septiembre de 2012 no es claramente inferior al que existía en

los meses previos a producirse el incremento del IVA reducido.

Dado que la aproximación visual no es suficiente para determinar si el aumento de tipos supuso un cambio de comportamiento en los hogares, es necesario contrastar si este diferencial en el porcentaje de hogares que consumían estos bienes y servicios en torno a los puntos de corte es estadísticamente significativo. Con este objeto, los efectos de la reforma del IVA analizada se estiman en un entorno de regresión en discontinuidad con puntos de corte acumulativos, donde los hogares reciben diferentes tratamientos o dosis de un mismo tratamiento para distintos rangos de la variable de ejecución, tiempo en este caso. Los hogares reciben el tratamiento 1 si $X_i < C_1$, en nuestro caso el tipo del 7 por 100 hasta junio 2010; el tratamiento 2 si $C_1 < X_i < C_2$, del 8 por 100 entre julio de 2010 hasta agosto de 2012; y así sucesivamente hasta el último valor de tratamiento en $X_i > C_j$, del 10 por 100 a partir de septiembre de 2012. Esto supone que las familias que consumen bienes y servicios en julio de 2010 tienen un incremento en el IVA de bienes afectados por tipo reducido de 1 punto (del 7 al 8 por 100), mientras que las familias que consumen bienes y servicios en septiembre de 2012 obtienen un aumento adicional del IVA de 2 puntos (del 8 al 10 por 100), es decir, un aumento en la dosis de tratamiento. Esto permite calcular dos impactos para los dos momentos de tiempo en que se produce el cambio en el tipo de gravamen del IVA, sobre la variable de interés que es el porcentaje de hogares que gastan en bienes debido a un aumento de tipos del IVA reducido: para el incremento de 1 punto en julio de 2010, el impacto es:

$$\begin{aligned} \text{impacto } p_{jul10,s} = & \lim_{+jul10} Phogbbss_{st} \\ & \text{porcentajes hogares consumen} \\ & \text{bbss grupo "s" después subida IVA} \\ & \text{(derecha jul 2010)} \\ - & \lim_{-jul10} Phogbbss_{st} \\ & \text{Porcentaje hogares consumen} \\ & \text{bbss grupo "s" antes subida IVA} \\ & \text{(izquierda jul 2010)} \end{aligned} \quad [2]$$

Es decir, se calcula el efecto que el aumento del 7 al 8 por 100 en el IVA tiene sobre el porcentaje de hogares que consumen ese tipo de bien «s» evaluando observaciones situadas alrededor de julio de 2010. De manera similar, para septiembre de 2012:

$$\begin{aligned} \text{impacto } p_{sep12,s} = & \lim_{+sep12} Phogbbss_{st} \\ & \text{porcentaje hogares consumen} \\ & \text{bbss grupo "s" después subida IVA} \\ & \text{(derecha sept 2012)} \\ - & \lim_{-sep12} Phogbbss_{st} \\ & \text{porcentaje hogares consumen} \\ & \text{bbss grupo "s" antes subida IVA} \\ & \text{(izquierda sept 2012)} \end{aligned} \quad [3]$$

Mediante la aproximación no paramétrica desarrollada por Calonico *et al.* (2014, 2015), se consideran dos tipos de tratamiento asociados al cambio normativo de julio de 2010 y septiembre de 2012. Estos valores de los tratamientos se definen como d_j , de modo que la variable de tratamiento viene dada por $D_i \in \{d_1, d_2, \dots, d_j\}$ y bajo condiciones de regularidad estándar el impacto alrededor de cada punto de corte « $C_j = \text{jul10, sep12}$ » es el dado por (9):

$$\begin{aligned} \text{impacto}_{j,s} = & E\{Y_i(d_j) - Y_i(d_{j-1}) | X_i = c_j\} = \\ = & \lim_{x \downarrow \text{jul10}} E[Y_i | X_i = x] - \lim_{x \uparrow \text{jul10}} E[Y_i | X_i = x] \end{aligned} \quad [4]$$

Donde X_i es el mes en que las familias realizan sus compras ($X_i = 1, \dots, 108$); Y_i es la variable de resultado donde se mide el impacto, que toma valor 1 si el hogar ha realizado gasto en ese grupo de bienes y 0 en caso contrario; y los puntos de corte en el tiempo donde se evalúa el impacto de la reforma son $c_1 = 43$, asociado con julio de 2010, y $c_2 = 69$ para septiembre de 2012 (10).

El cuadro n.º 1 recoge el estimador no paramétrico del impacto sobre el porcentaje de hogares

CUADRO N.º 1
ESTIMACIÓN DEL IMPACTO SOBRE EL PORCENTAJE DE HOGARES QUE GASTAN EN BIENES AFECTADOS POR UN INCREMENTO DE LOS TIPOS DEL IVA REDUCIDO

	REFORMA 2010	REFORMA 2012
Grupo de bienes	Estimación impacto (p-valor)	Estimación impacto (p-valor)
Alimentos	-0,003 (0,634)	-0,008* (0,031)
Bebidas	-0,004 (0,818)	-0,019* (0,022)
Restauración	-0,010 (0,456)	-0,018 (0,296)
Hostelería	0,032* (0,086)	0,014 (0,857)
Transporte	-0,007 (0,825)	0,005 (0,904)

* Estimador es estadísticamente significativo al 10 por 100.

CUADRO N.º 2

P-VALOR DE LA ESTIMACIÓN DEL IMPACTO DE LA SUBIDA DE IVA REDUCIDO EN JULIO DE 2008 (CONTRASTE DE FALSIFICACIÓN)

	ALIMENTOS	BEBIDAS	RESTAURACIÓN	HOSTELERÍA	TRANSPORTE
Reforma de julio de 2008	0,26	0,68	0,18	0,98	0,46

que gastan en bienes y servicios que supone el incremento del tipo del IVA en las dos fechas consideradas y entre paréntesis el *p*-valor del contraste de significatividad. Los resultados muestran que el incremento del tipo de gravamen del IVA de 2 puntos en 2012 redujo muy ligeramente el número de familias que consumían bienes y servicios afectados por el IVA reducido, marginalmente en alimentos, y casi 2 puntos probabilísticos en bebidas (*p*-valor menor de 0,1) respecto a una situación hipotética en que no se hubieran subido los tipos de gravamen. Es decir, un aumento del 8 al 10 por 100 en el tipo de gravamen del IVA supuso un ligero descenso en el número de hogares que consumen bienes y servicios con IVA reducido. Sin embargo, los resultados obtenidos para la reforma de 2010 no son significativos, salvo en hostelería, donde aumenta el número de familias que consumen estos servicios.

La validez de la estimación del impacto de una intervención pública empleando el enfoque de regresión en discontinuidad depende del cumplimiento de un conjunto de supuestos. Por ello, se lleva a cabo un análisis de robustez de los resultados obtenidos mediante los contrastes de falsificación, ausencia de manipulación y equilibrado de muestras. El contraste de falsificación (o placebo) analiza si la discontinuidad explotada para obtener los impactos causales del incremento de tipo del IVA fue originada por el cambio normativo de 2010 y 2012. Para comprobar este hecho se propone utilizar un punto de corte falso, por ejemplo, julio de 2008, y estimar el impacto sobre el porcentaje de hogares que gastan en bienes afectados por los tipos del IVA reducido alrededor de esa fecha. Los resultados dados en el cuadro n.º 2 muestran que no se producen cambios estadísticamente significativos en los niveles de consumo.

El contraste de ausencia de manipulación determina si los hogares pueden decidir su posición en torno al umbral, julio de 2010 y septiembre de 2012 respectivamente. Si los hogares no pueden manipular su posición en torno a la frontera, se debería observar un número parecido de observaciones con anterioridad y con posterioridad a la fecha

de reforma. Por el contrario, si las unidades tuvieran el poder de afectar su puntuación (valor del eje de las *x* en los gráficos) y saben que el tratamiento es beneficioso, deberíamos esperar más personas justo por debajo del límite (donde el IVA es menor) que por encima. El contraste de McCrary (2008), con un *p*-valor de 0,342 refleja la ausencia de manipulación en torno a los dos valores críticos.

Finalmente, el contraste de equilibrado de muestras analiza si hay diferencias significativas en otras características de los individuos alrededor de los meses de julio de 2010 y septiembre de 2012. Si esto ocurriera existiría un problema de identifi-

CUADRO N.º 3

P-VALORES DEL CONTRASTE DE DIFERENCIA DE MEDIAS DE CARACTERÍSTICAS DE LOS HOGARES EN TORNO A LOS PUNTOS DE CORTE DE 2010 Y 2012

CARACTERÍSTICAS DEL HOGAR	REFORMA 2010	REFORMA 2012
Cabeza familia hombre (var. dicotómica [1,0])	0,28	0,59
Hogar en municipio pequeño (var. dicotómica [1,0])	0,29	0,18
Número miembros del hogar	0,53	0,08
Edad del cabeza de familia	0,68	0,11
Ingresos del hogar	0,41	0,96
Hogar reside en CC. AA. (var. dicotómica [1,0])		
Andalucía	0,76	0,11
Aragón	0,94	0,77
Asturias	0,87	0,68
Baleares	0,59	0,57
Cantabria	0,23	0,57
Castilla y León	0,13	0,72
Castilla-La Mancha	0,33	0,39
Cataluña	0,22	0,31
Comunidad Valenciana	0,90	0,86
Extremadura	0,37	0,95
Galicia	0,71	0,63
Madrid	0,61	0,82
Murcia	0,66	0,89
Navarra	0,11	0,89
País Vasco	0,34	0,99
La Rioja	0,85	0,64

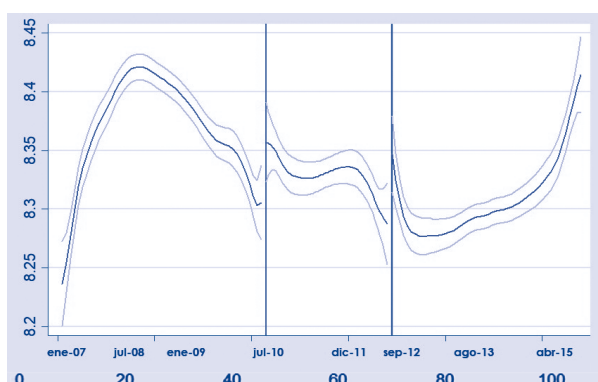
cación debido a que no sería posible determinar si la diferencia en el porcentaje de hogares que consumen bienes y servicios sujetos al IVA reducido es consecuencia de la reforma o de las diferencias en las características entre los dos grupos. En el cuadro n.º 3 se muestra el p -valor del contraste de diferencia de medias para distintas características de los hogares alrededor de los dos puntos de corte considerados en 2010 y 2012, observándose que no existen diferencias significativas.

Todos estos resultados anteriores reflejan la validez de la metodología empleada de regresión en discontinuidad. A continuación, se estima el impacto de las reformas en el margen intensivo, analizando, para aquellos que consumen estos bienes, la evolución del gasto medio del hogar a lo largo del tiempo.

2. Efecto sobre el gasto medio de los hogares en bienes y servicios (margen intensivo)

De forma similar a la estimación anterior, se calcula inicialmente el gasto promedio que realizan los hogares en bienes afectados por el IVA reducido a lo largo del tiempo, mediante la regresión dada por $E(\text{Gasto}_{hst} | \text{mes} = x)$ empleando técnicas no paramétricas. El gráfico 5 muestra el gasto promedio de los hogares en bienes afectados por tipos reducidos, de forma mensual, en el período de enero 2007 a diciembre 2015, prestando especial atención a los momentos en los que se produce el cambio de tipos del IVA, en julio de 2010 del 7 al 8 por 100, y en septiembre de 2012, del 8 al 10 por 100.

GRÁFICO 5
LOG DEL GASTO MEDIO EN BIENES Y SERVICIOS
CON IVA REDUCIDO. ENE/2007 – DIC/2015



El gráfico muestra el estimador no paramétrico, empleando polinomios locales de tercer orden, del gasto promedio de los hogares en los bienes afectados por el IVA reducido calculado en cada uno de los tres tramos relevantes: antes de julio de 2010, entre julio de 2010 hasta septiembre de 2012, y posteriormente, observándose que los valores cercanos a las dos fechas frontera estudiadas hay cambios en los patrones de gasto de los hogares, aunque no resulta claro que sean estadísticamente significativos dado que los intervalos de confianza por ambos lados de los valores críticos de julio de 2010 y septiembre de 2012 se solapan. Para determinar la existencia del impacto en el año 2010 y 2012 respectivamente se estiman las siguientes ecuaciones:

$$\begin{aligned} \text{impacto } g_{\text{jul10}} = & \lim_{+ \text{ jul10}} \log \text{gasto}_t - \\ & \text{gasto medio hogares consumen} \\ & \text{bienes con IVA reducido depues subida IVA} \\ & \text{(derecha jul 2010)} \\ & - \lim_{- \text{ jul10}} \log \text{gasto}_t \\ & \text{gasto medio hogares consumen} \\ & \text{bienes con IVA reducido antes subida IVA} \\ & \text{(izquierda jul 2010)} \end{aligned} \quad [5]$$

$$\begin{aligned} \text{impacto } g_{\text{sep12}} = & \lim_{+ \text{ sep12}} \log \text{gasto}_t - \\ & \text{gasto medio hogares consumen} \\ & \text{bienes con IVA reducido depues subida IVA} \\ & \text{(derecha sept 2012)} \\ & - \lim_{- \text{ sep12}} \log \text{gasto}_t \\ & \text{gasto medio hogares consumen} \\ & \text{bienes con IVA reducido antes subida IVA} \\ & \text{(izquierda sept 2012)} \end{aligned} \quad [6]$$

Utilizando el método de estimación no paramétrica de Calonico *et al.* (2014, 2015), los resultados muestran que el incremento del IVA en julio de 2010 no tiene efectos significativos en el gasto medio de los hogares, mientras que el aumento de 2 puntos de septiembre de 2012 redujo el gasto medio de los hogares en bienes y servicios (p -valor menor de 0,1) en casi 1 punto logarítmico, respecto a una situación hipotética en que no se hubieran subido los tipos de gravamen.

De manera similar al análisis del impacto en el margen extensivo, a continuación se muestra el

CUADRO N.º 4

ESTIMACIÓN DEL IMPACTO EN EL GASTO MEDIO DE HOGARES POR SUBIDA DE TIPOS DEL IVA

	IMPACTO	P-VALOR
Reforma 2010	-0,021	0,96
Reforma 2012	-0,096*	0,01

GRÁFICO 6
(LOG) GASTO MEDIO EN ALIMENTOS (IZQDA.) Y BEBIDAS NO ALCOHOL (DCHA.) ENE/07-DIC/15

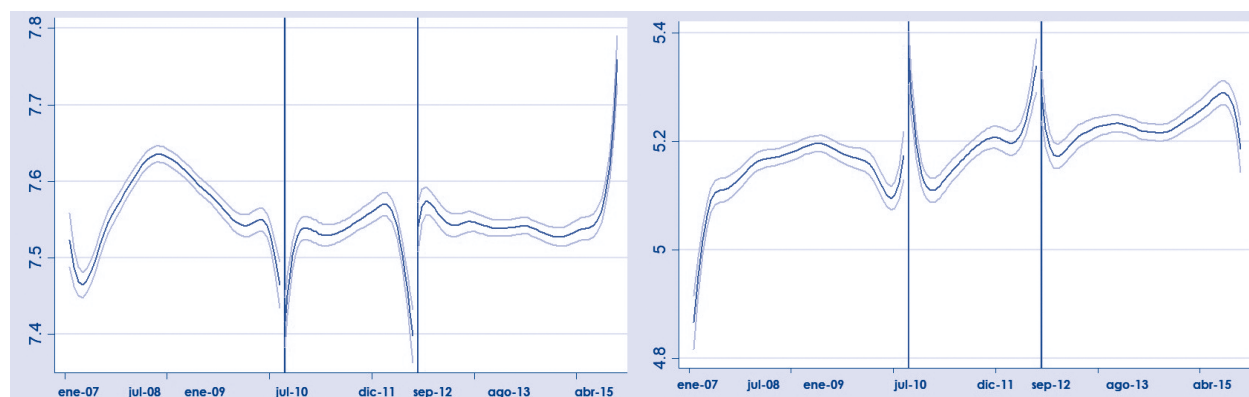
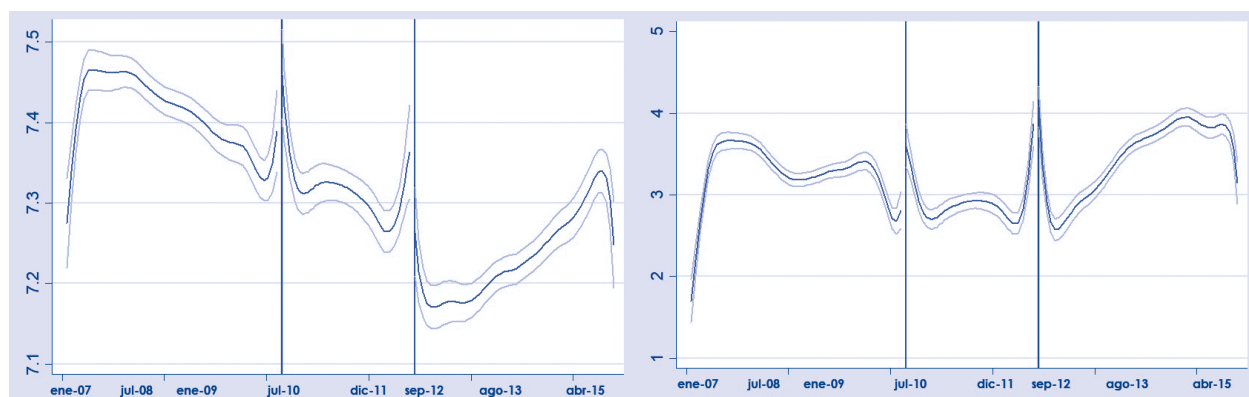


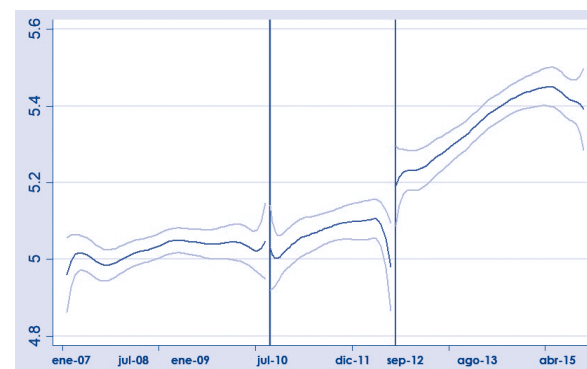
GRÁFICO 7
(LOG) GASTO MENSUAL EN RESTAURANTES (IZQDA.) Y HOSTELERÍA (DCHA.) ENE/07-DIC/15



efecto en el margen intensivo para diferentes tipos de bienes y servicios, medido en el gasto medio de los hogares, $E(\text{Gasto}_{hst} | \text{mes} = x)$, donde Gasto_{hst} indica el gasto de un hogar «h» en el bien «s» en el mes «t». Esta aproximación permite aislar el efecto sobre cada grupo de bienes o servicios excluyendo aquellos hogares que pasaron a consumir cero. En términos gráficos el impacto que la subida de tipos del IVA reducido ha tenido en el gasto de los hogares viene representado en los gráficos 6 a 8.

De acuerdo a los valores del $E(\text{Gasto}_{hst} | \text{mes} = x)$ a la derecha e izquierda de julio de 2010 y septiembre de 2012, parece producirse una caída en el consumo de alimentos en 2010, y de bebidas, restauración y hostelería en 2012, mientras que, por el lado contrario, se observa un incremento del gasto medio

GRÁFICO 8
(LOG) GASTO MEDIO EN TRANSPORTE. ENE/07-DIC/15



CUADRO N.º 5

ESTIMACIÓN DEL IMPACTO EN EL GASTO MEDIO DE HOGARES
POR LA SUBIDA DE TIPOS DEL IVA

Grupo de Bienes	REFORMA 2010	REFORMA 2012
	Estimación impacto (p-valor)	Estimación impacto (p-valor)
Alimentos	-0,023 (0,482)	-0,004 (0,382)
Bebidas	-0,169* (0,021)	-0,091* (0,013)
Restauración	0,051 (0,253)	-0,177* (0,002)
Hostelería	0,249 (0,202)	-1,199* (0,0009)
Transporte	-0,353* (0,029)	0,131* (0,068)

en alimentos y transporte en 2012, siendo el resto de valores no especialmente relevantes. De manera similar al gasto total, el cuadro n.º 5 muestra el impacto que el incremento de tipos en el IVA reducido en esas dos fechas ha supuesto en el gasto de los hogares aplicando la fórmula dada en [5] y [6].

Hay impactos estadísticamente relevantes, sobre todo en la reforma impositiva de 2012, con un descenso en el gasto medio de los hogares en bebidas no alcohólicas, restauración y hostelería, mientras que el gasto en transporte aumenta. Por lo que respecta al gasto en alimentos no se observan cambios significativos. El incremento de 1 punto de julio de 2010 solo impacta en una reducción del gasto en bebidas y transporte, siendo ineficaz en el resto de los casos. Resultan interesantes los resultados obtenidos en alimentación, dado que se trata del único

GRÁFICO 9

IMPACTO SOBRE PORCENTAJE DE HOGARES CON GASTO EN ALIMENTOS, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DE IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

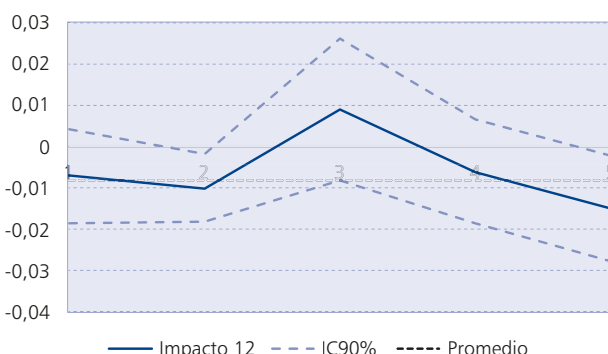
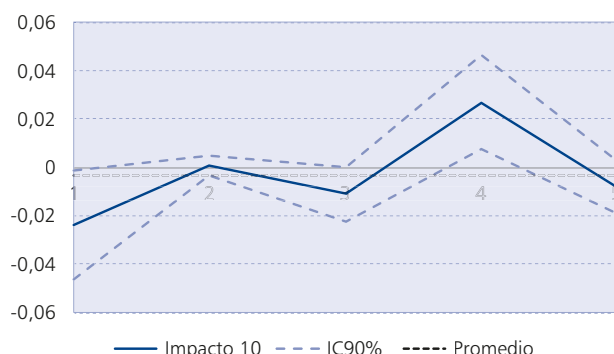


GRÁFICO 10

IMPACTO SOBRE PORCENTAJE DE HOGARES CON GASTO EN BEBIDAS NO ALCOHÓLICAS, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

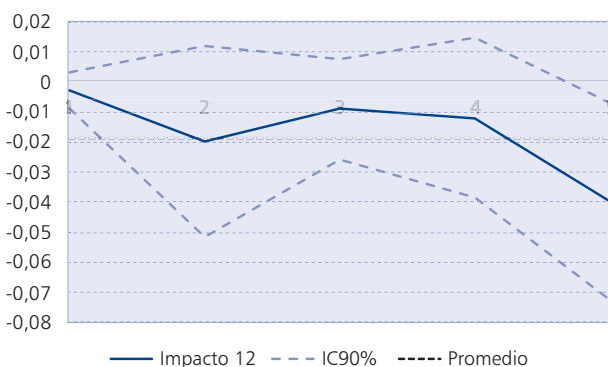
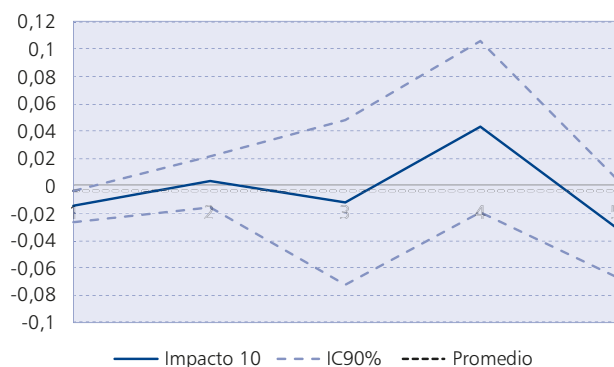


GRÁFICO 11

IMPACTO SOBRE PORCENTAJE DE HOGARES CON GASTO EN RESTAURACIÓN, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

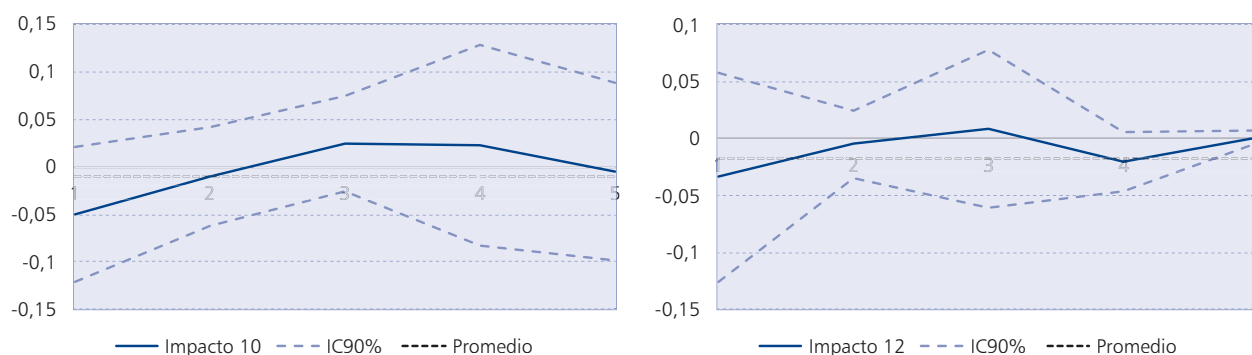


GRÁFICO 12

IMPACTO SOBRE PORCENTAJE DE HOGARES CON GASTO EN HOSTELERÍA, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

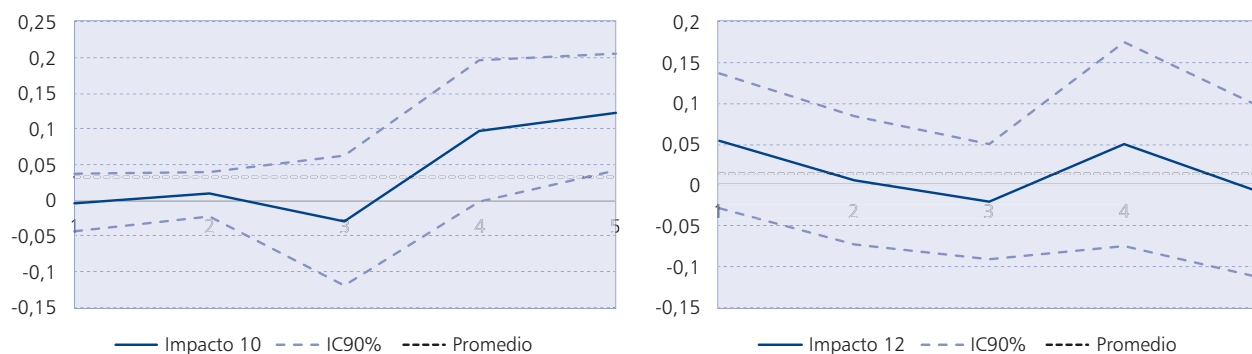
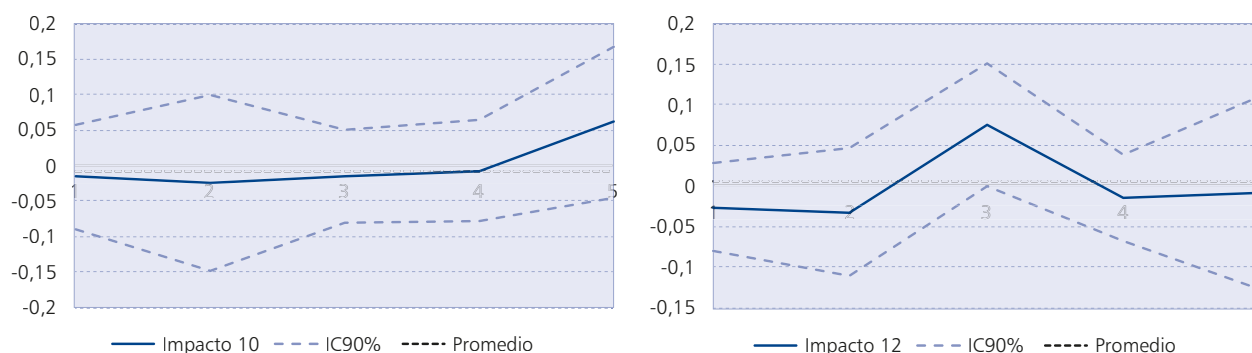


GRÁFICO 13

IMPACTO SOBRE PORCENTAJE DE HOGARES CON GASTO EN TRANSPORTE, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)



Nota: La línea horizontal indica el impacto promedio, la línea continua negra es el impacto para cada cuantil de renta y las líneas discontinuas negras son los intervalos al 90 por 100 del impacto.

de los grupos de bienes donde no hay respuesta clara por parte de los hogares, confirmando que se trata de un bien muy inelástico.

3. Análisis de heterogeneidad: impacto por tipología de hogar

En esta última sección se analizan los efectos heterogéneos de la subida de gravamen del IVA reducido, estudiando qué tipo de familias son las afectadas principalmente por la reforma impositiva. Para ello, se agrupa a los hogares de acuerdo con el quintil de renta del hogar suministrada en la EPF. A continuación, se evalúa el impacto de subida de tipos reducidos para cada grupo de hogares y por tipo de bienes en el margen extensivo, viendo el cambio en el número de hogares que dejaron de consumir ese bien o servicio, y con posterioridad se presentaran los resultados en el margen intensivo.

Análisis de heterogeneidad en el margen extensivo

Aplicando la estimación dado en [2] para las dos reformas analizadas, y la diferente tipología de hogares agrupados por quintiles de renta, los gráficos números 9 a 13 muestran el impacto sobre el porcentaje de hogares que consumían diferentes tipos de bienes o servicios afectados por el cambio de gravamen del IVA reducido. Cada gráfico muestra, por un lado, el impacto promedio para toda la población, mostrado en el cuadro n.º 1, que viene representado por la línea horizontal discontinua, mientras que la línea continua negra indica el impacto para cada quintil de renta, junto con sus intervalos de confianza al 90 por 100 dados por las líneas negras discontinuas.

Observando aquellas zonas donde los intervalos de confianza no contienen el 0 se puede determinar aquellos niveles de renta donde el porcentaje de hogares que consumían ese bien o servicio cambió

GRÁFICO 14

IMPACTO SOBRE PORCENTAJE EN GASTO EN ALIMENTOS, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

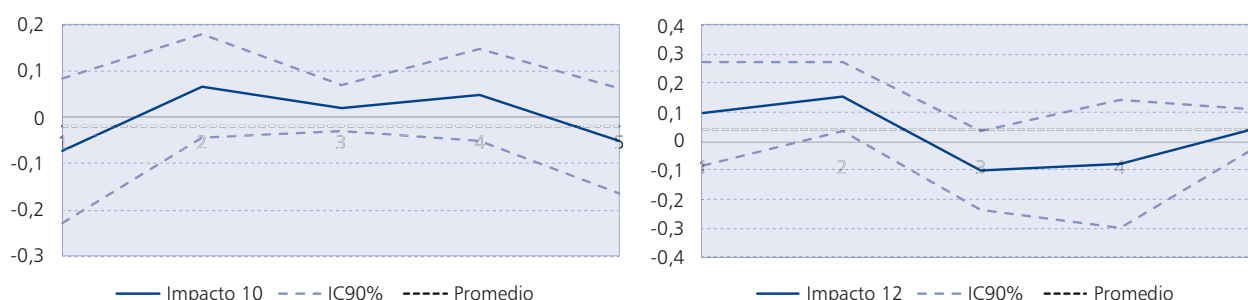


GRÁFICO 15

IMPACTO SOBRE PORCENTAJE EN GASTO EN BEBIDAS, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

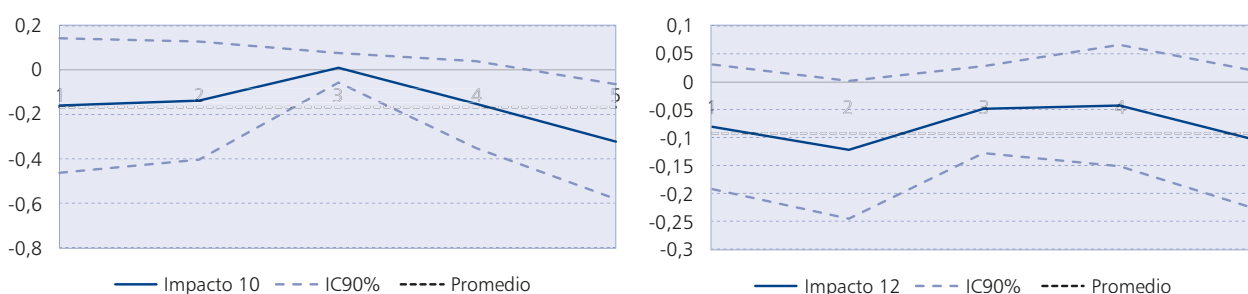


GRÁFICO 16

IMPACTO SOBRE PORCENTAJE EN GASTO EN RESTAURACIÓN, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

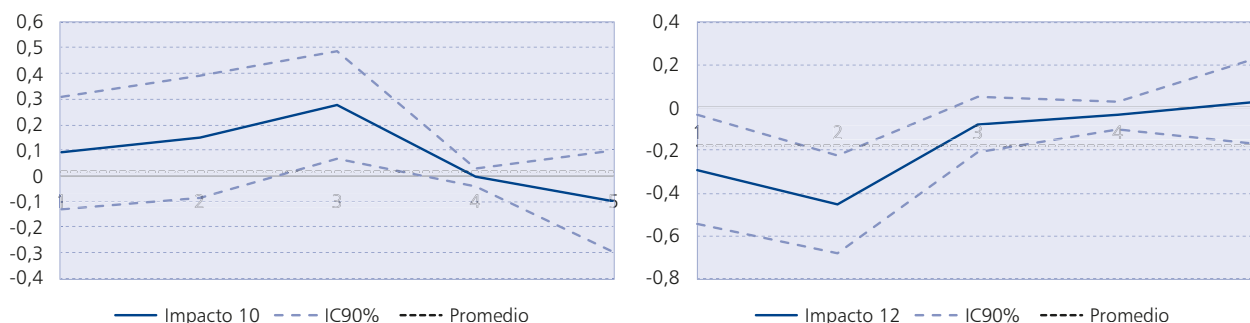


GRÁFICO 17

IMPACTO SOBRE PORCENTAJE EN GASTO EN HOSTELERÍA, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)

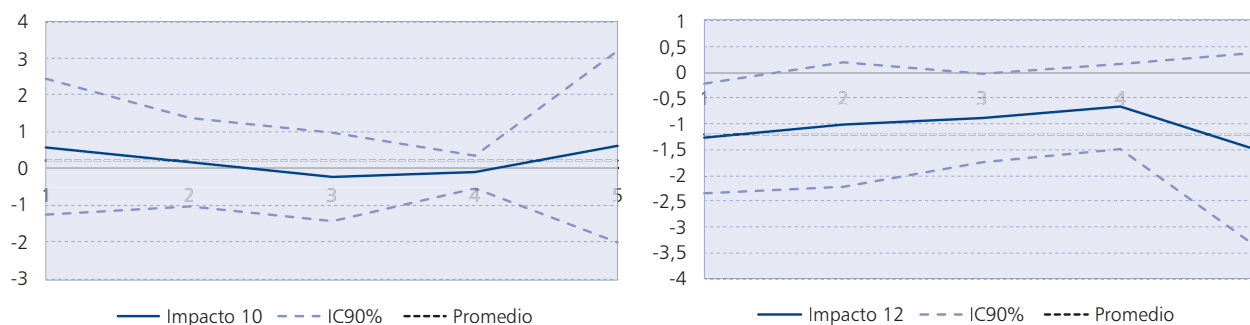
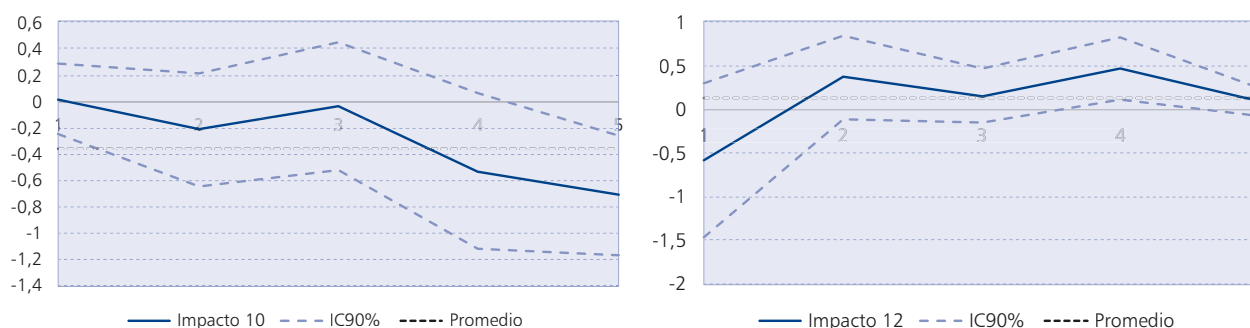


GRÁFICO 18

IMPACTO SOBRE PORCENTAJE EN GASTO EN TRANSPORTE, POR QUINTILES RENTA HOGAR, DE LA SUBIDA DE TIPOS DEL IVA. REF. 10 (IZQDA.) REF. 12 (DCHA.)



significativamente debido a una modificación del gravamen del IVA reducido. Destaca que la reforma del IVA de 2010 afecta principalmente a rentas bajas, con reducciones en el número de hogares que gastan en alimentos y bebidas no alcohólicas en torno a 2 puntos porcentuales, mientras que, por el contrario, la reforma de 2012 impacta más en rentas altas.

Análisis de heterogeneidad en el margen intensivo

A continuación, se ofrece el efecto de la reforma del IVA sobre el gasto medio de los hogares en diferentes tipos de bienes y servicios dependiendo de su nivel de renta. Como el caso del margen extensivo, las figuras muestran el impacto promedio en el gasto medio realizado en ese bien para toda la población dado en el cuadro n.º 5, que viene representado por la línea horizontal discontinua, mientras que la línea continua negra indica el impacto sobre el gasto en ese bien para cada quintil de renta, mostrándose los intervalos de confianza al 90 por 100 por las líneas negras discontinuas.

Se observa que el cambio de gravamen no tiene efectos significativos sobre el gasto medio de los hogares en alimentos, independientemente del nivel de ingresos del hogar, reflejando la baja elasticidad-renta existente para este tipo de bienes, mientras que para bebidas no alcohólicas y restauración hay un efecto en la zona media de la distribución. Finalmente, las rentas bajas reducen su gasto en hostelería tras la subida del IVA como refleja la gráfica de la reforma de 2012, y en el gasto en transporte, la subida de tipos de 2010 influye en una reducción del gasto medio de los hogares con rentas altas.

IV. CONCLUSIONES

Este trabajo analiza el impacto que un incremento del tipo de gravamen en el IVA reducido en los años 2010 y 2012 de 1 y 2 puntos respectivamente ha tenido sobre el gasto de los hogares en España, empleando una aproximación de evaluación con contrafactuales, y usando la metodología de regresión en discontinuidad para la estimación. Con datos mensualizados de la EPF para el período 2007-2015, se evalúa tanto el cambio en el número de hogares que consumen bienes o servicios afectados por el IVA reducido –margen extensivo– como el efecto sobre el gasto medio de los hogares –margen intensivo–, desagregando los análisis para diferentes agrupaciones de bienes y servicios

así como tipologías de hogares. Los resultados muestran, respecto al margen extensivo, que la subida de dos puntos de 2012 reduce ligeramente el número de hogares que gastan en alimentos y bebidas, mientras que el porcentaje de hogares que realizan gastos en hostelería aumenta en 2010. En el margen intensivo, la reforma de 2012 afecta negativamente al gasto medio de los hogares en este tipo de bienes y servicios, influyendo especialmente en una reducción en el gasto en bebidas no alcohólicas, restauración y hostelería. Por lo que respecta al análisis de heterogeneidad dependiendo del nivel de ingresos del hogar, el resultado más relevante es el efecto negativo de la subida de gravamen de 2010 en los hogares de renta baja, con una reducción del número de consumidores en alimentos y bebidas, mientras que la reforma de 2012 afectó negativamente a rentas altas en estos tipos de bienes. Para el gasto medio de los hogares, la subida de 2010 afecta principalmente a rentas altas en el gasto en bebidas no alcohólicas y transporte, mientras que la de 2012 influye, sobre todo, en hogares con rentas bajas con una reducción del gasto medio en diferentes bienes y servicios.

NOTAS

(1) Entre otros trabajos, consultar SANZ y ROMERO (2012a) (2012b) o GARCÍA-ENRÍQUEZ y ECHEVARRÍA (2018).

(2) Aunque el documento de la AIREF (2020) presenta ciertas similitudes con la aproximación de este trabajo, la propuesta presentada en este documento tiene claras mejoras, como: a) la evaluación conjunta de las dos reformas de 2010 y 2012; b) proponer un método de estimación consistente para evaluar simultáneamente ambas reformas; c) el análisis del efecto tanto en el margen extensivo como intensivo; y d) un análisis de robustez de las estimaciones.

(3) Estos segundos autores también proponen un modelo de generaciones solapadas como alternativa de estimación.

(4) En este sentido existen argumentos a favor de la utilización de la renta permanente en lugar de la renta corriente en el análisis distributivo de los impuestos (POTERBA, 1991).

(5) En la mayoría de estos trabajos, para el nivel de progresividad de la reforma analizada se suele presentar la distribución de tipos medios efectivos, índices de KAKWANI (1977) y SUITS (1977) o la elasticidad impositiva propuesta por MUSGRAVE y THIN (1948).

(6) Dentro de las herramientas para simular reformas de impuestos indirectos destacan dos aplicaciones de simulación como el Simulador de Imposición Indirecta del Instituto de Estudios Fiscales y el FUNCAS-Sindi de FUNCAS, con opciones de simulación tanto con, como sin, comportamiento a partir de la estimación de modelos tipo AIDS. Para más detalles consultar Instituto de Estudios Fiscales (2011) y SANZ *et al.* (2003; 2013).

(7) Existe una amplia literatura que también analiza el carácter regresivo de los impuestos indirectos en España. Para más detalles se pueden consultar, entre otros, los trabajos de GONZÁLEZ-PÁRAMO y SALAS (1991); MAYO y SALAS (1994); SANZ, ROMERO y CASTAÑER (2013); SANZ y ROMERO (2012a, 2012b); y ROMERO, SANZ y CASTAÑER (2013).

(8) Es importante destacar que a la hora de realizar un análisis sobre los efectos de cambios en un impuesto indirecto, existen una serie de factores que influyen a la hora de determinar cómo una variación en el tipo de gravamen del IVA incide sobre el consumo de bienes y servicios de los hogares: la variación impositiva, el efecto de la reducción del impuesto sobre los precios –analizando la traslación de los productores al precio final– y la elasticidad-precio de la demanda. En nuestro caso particular, se realiza una evaluación, en la que se considera de forma conjunta estos factores mencionados previamente, no realizando ninguna desagregación por tipología de factores. Para más información sobre la traslación de cambios impositivos en los precios, se pueden consultar los trabajos de GAARDER (2016) y NIPERS *et al.* (2019) para alimentos; KOSONEN (2015) para peluquería; BENZARTI *et al.* (2020) en restaurantes; CARBONNIER (2007) en venta de coches; o MORAL-ARCE y GÓMEZ-ANTONIO (2020) para bienes culturales.

(9) A diferencia del caso con múltiples puntos de corte no acumulativos, la población no se encuentra particionada, de tal modo que cada observación se puede usar para estimar dos diferentes efectos del tratamiento (pero contiguas en la dimensión de la puntuación, en nuestro caso el mes de consumo). Por ejemplo, unidades que reciben dosis de tratamiento d_j se usan como «tratados» (es decir, por encima del límite c_j – julio de 2010) al estimar el impacto en el primer corte, que denominamos α_j , pero también se pueden usar como «controles» al estimar el impacto en el siguiente punto de corte, α_{j+1} (es decir, debajo del límite c_{j+1} – septiembre de 2012). Es posible que este hecho afecte a la independencia de los estimadores empleados en los diferentes puntos de corte, aunque este problema de dependencia desaparece asintóticamente mientras la amplitud de banda empleada alrededor de cada punto de corte disminuyen con el tamaño de la muestra. Por otra parte, estos anchos de banda se pueden elegir para que no se superpongan y de este modo garantizar que las observaciones se utilicen una única vez.

(10) Otra posibilidad es obtener un estimador de α_j mediante la aproximación clásica de Hahn *et al.* (2001) y PORTER (2003) utilizando un enfoque no paramétrico de polinomios locales basados en *kernels* a ambos lados del límite c_j . Estos estimadores de regresión no paramétricos son particularmente adecuados para la inferencia en el diseño de regresión en discontinuidad debido a sus buenas propiedades en el límite del soporte de la regresión. Para más detalles ver FAN y GIJBELS (1996).

BIBLIOGRAFÍA

- AIReF (2020). *Estudio de Beneficios Fiscales. Evaluación de gasto público 2019*. Autoridad Independiente de Responsabilidad Fiscal (AIReF).
- ARTÉS, J., BOTELLO, A. M. y SÁNCHEZ-FUENTES, A. (2019) Tax Reforms and Google Searches: The Case of Spanish VAT Reforms during the Great Recession. *SERIEs*, vol. 10, n.º 3-4, pp. 321-336.
- ASENSIO, J., MATAS, A. y RAYMOND, J. L. (2003). Petrol Expenditure and Redistributive Effects of Its Taxation in Spain. *Transportation Research: Part A: Policy and Practice*, 37, pp. 49-69.
- ATECA-AMESTOY, V., GARDEAZÁBAL, J. y UGIDOS, A. (2020). On the Response of Household Expenditure on Cinema and Performing Arts to Changes in Indirect Taxation: A Natural Experiment in Spain. *Journal of Cultural Economics*, 44(2), pp. 213-253.
- BAJO-BUENESTADO, R. (2017). Does a Gas Tax Cut Reduce Retail Fuel Prices? *Journal of Transport Economics and Policy*, 51(3), pp. 157-72.
- BELLO, A. y CONTIN-PILART, I. (2012). Taxes, Cost and Demand Shifters as Determinants in the Regional Gasoline Price Formation Process: Evidence from Spain. *Energy Policy*, 48(1), pp. 439-48.
- BENZARTI, Y., CARLONI, D., HARJU, J. y KOSONEN, T. (2020) What goes up may not come down, asymmetric incidence of value added taxes. *Journal of political economy*, 128, pp. 4438-4474.
- CARBONNIER, C. (2007). Who pays sales taxes? Evidence from French VAT reforms, 1987–1999. *Journal of Public Economics*, issue 5-6, pp. 1219-1229.
- CONESA, J. C., DÍAZ-GIMÉNEZ, J., DÍAZ-SAAVEDRA, J. y PUJOAN-MAS, J. (2010). La subida del impuesto sobre el valor añadido en España: demasiado cara y demasiado pronto. *Documento de Trabajo de FEDEA*, 2010-06.
- CALONICO, S., CATTANEO, M. D., FARRELL, M. H. y TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82, pp. 2295-2326.
- CALONICO, S., CATTANEO, M. D., FARRELL, M. H. y TITIUNIK, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110, pp. 1753-1769.
- ECHEVARRÍA, C. A. y GARCÍA-ENRÍQUEZ, J. (2014). 2012 VAT reform in Spain: The food and non-alcoholic beverages case. Welfare consequences. *Documento de Trabajo*. Universidad del País Vasco UPV.
- FAN, J. y GIJBELS, I. (1996). Local polynomial modelling and its applications. *Monographs on Statistics and Applied Probability* (Series) 66. CRC Press.
- FERNÁNDEZ DE CÓRDOBA, G. y TORRES, J. L. (2010). El aumento del IVA en España: una cuantificación anticipada de sus efectos. *Revista de Economía Aplicada*, 18(53), pp. 163-183.
- GAARDER, I. (2016). Incidence and distributional effects of value added taxes. *The Economic Journal*, 129(618), pp. 853-876.
- GARCÍA-ENRÍQUEZ, J. y ECHEVARRÍA, C. A. (2018). Demand for Culture in Spain and the 2012 VAT Rise. *Journal of Cultural Economics*, 42(3), pp. 469-506.
- GONZÁLEZ-PÁRAMO, J. M. y SALAS, R. (1991) Incidencia redistributiva del IVA. *Hacienda Pública Española*, 119, pp. 89-104.
- HAHN, J., TODD, P. y VAN DER KLAUW, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), pp. 201-209.
- INSTITUTO DE ESTADÍSTICA (2010) *Encuesta de Presupuestos Familiares Base 2006*. Madrid.
- INSTITUTO DE ESTUDIOS FISCALES (2011). Herramientas de simulación impositiva del Instituto de Estudios Fiscales. *Documentos de Trabajo*. IEF, Ministerio de Hacienda.
- IZQUIERDO, J. F., MELGUIZO, A. y TAGUAS, D. (2001) Imposición y precios de consumo. *Papeles de Economía Española*, 87, pp. 199-220.

- KAKWANI, N. C. (1977). Measurement of tax progressivity: an international comparison. *Economic Journal*, 87, pp. 71-80.
- KEHOE, T. J. (1989) Un análisis de equilibrio general de la reforma fiscal de 1986 en España. *Investigaciones Económicas*, 13(3), pp. 337-85.
- KOSONEN, T. (2015). More and cheaper haircuts after VAT cut? On the efficiency and incidence of service sector consumption taxes. *Journal of Public Economics*, 131, pp. 87-100.
- LABEAGA, J. M. y LÓPEZ, A. (1994) Estimation of the welfare effects of indirect tax changes on Spanish households: an analysis of the 1992 VAT reform. *Investigaciones Económicas*, 18, pp. 289-311.
- LABEAGA, J. M. y LÓPEZ, A. (1996). Flexible Demand System Estimation and the Revenue and Welfare Effects of the 1995 Vat Reform on Spanish Households. *Revista Española de Economía*, 13(2), pp. 181-197.
- LÓPEZ, E., GONZÁLEZ, M. A. y FELLINGER, E. (2003). La demanda regional de bebidas alcohólicas en España. Impacto de un incremento de los impuestos especiales. *Economía Agraria y Recursos Naturales*, 3(1), pp. 123-138.
- LÓPEZ-LABORDA, J., MARÍN, C. y ONRUBIA, J. (2008). Tipo reducido, superreducido y exenciones en el IVA: una estimación de sus efectos recaudatorios y distributivos a partir de las encuestas de hogares. *Estudios de Economía Española*, 2018-23. FEDEA.
- MAYO, R. y SALA, R. (1994). Incidencia redistributiva del IVA. Tipos efectivos declarados. *Hacienda Pública Española*, 128, pp. 33-74.
- MARTÍNEZ-VÁZQUEZ, J. y BIRD, R. M. (2010). El impuesto sobre el valor añadido: ¿hacia adelante y hacia arriba? *Papeles de Economía Española*, 125-126, pp. 223-261.
- MCCRARY, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), pp 698-714.
- METCALF, G. E. (1994). Life cycle versus annual perspectives on the incidence of a value added tax. En A. R. MOFFIT (ed.), *Tax policy and the economy*, 8, pp. 45-64.
- MORAL-ARCE, I. y GÓMEZ-ANTONIO, M. (2020). Una evaluación sobre los efectos de una reducción del tipo de gravamen del IVA para los bienes y servicios culturales. *Papeles de Trabajo IEF*, 4/2020. Instituto de Estudios Fiscales.
- MUSGRAVE, R. A. y THIN, T. (1958). Income tax progression, 1929-1048. *Journal of Political Economy*, 56, pp. 498-514.
- NIPERS, A., UPÍTE, I., PILVERE, I., STALGIENE, A., y VIIRA, A. (2013). Effect of VAT rate reduction for fruits and vegetables on prices in Latvia: ex-post analysis. *Agrarteadus Journal of Agricultural Science*, pp 25-31.
- PORTER, J. (2003). Estimation in the Regression Discontinuity Model. *Working Paper*. University of Wisconsin, pp. 2297-2300.
- POTERBA, J. M. (1991). Is the gasoline tax regressive? En D. BRADFORD (ed.), *Tax policy and the economy*, 5, pp. 145-164. MIT Press.
- PRIETO, J. y ROMERO, D. (1977). Is a tax cut on cultural goods consumption actually desirable? A microsimulation analysis applied to Spain. *Fiscal Studies*, 26, pp. 549-575.
- ROMERO, D. (2014) Los impuestos indirectos en España: diagnosis y discusión de propuestas factibles de reforma. *Papeles de Economía Española*, 139, pp. 39-55.
- ROMERO, D. (2017). Progresividad y capacidad redistributiva del IVA e impuestos especiales en el periodo 2007-2015. *Papeles de Economía Española*, 154, pp. 200-215.
- ROMERO, D., SANZ, J. F. y CASTAÑER, J. M. (2013). Sobre la regresividad de la imposición indirecta en España en tiempos de crisis: un análisis con microdatos de hogares. *Papeles de Economía Española*, 135, pp 172-183.
- SANZ, J. F., ROMERO, D., CASTAÑER, J. M., PRIETO, J. y FERNÁNDEZ, F. J. (2003). *Microsimulación y comportamiento económico en el análisis de reformas de imposición indirecta. El Simulador de Imposición Indirecta del Instituto de Estudios Fiscales*. Madrid: Instituto de Estudios Fiscales.
- SANZ, J. F. y ROMERO, D. (2012a). Efectos recaudatorios y distributivos en 2012 de algunas opciones de reforma del IVA. Impacto sobre el sector de los hogares. *Cuadernos de Información Económica*, 228, pp. 1-5.
- SANZ, J. F. y ROMERO, D. (2012b). Efectos recaudatorios de la reforma del IVA de julio de 2012. *Cuadernos de Información Económica*, 229, pp. 1-4.
- SANZ, J. F. y ROMERO, D. (2012c) Evaluación del impacto recaudatorio de las reformas del IRPF e IVA de 2012 en un contexto de crisis económica. *Papeles de Economía Española*, 133, pp. 39-48.
- SANZ, J. F., ROMERO, D. y CASTAÑER, J. M. (2013). *El análisis de los impuestos indirectos a partir de las Encuestas de Presupuestos Familiares. El simulador de Impuestos Indirectos de la Fundación de las Cajas de Ahorros (FUNCASIndi)*. Madrid: Fundación de las Cajas de Ahorros.
- SERRANO, A. y OLIVA, N. (2011). Efectos distributivos de la reforma 2010 del IVA en España: Otras propuestas alternativas. *Estudios de Economía Política*, 19, pp. 67-86.
- SUITS, D. B. (1977). Measurement of tax progressivity. *American Economic Review*, 67, pp. 747-752.

LOS PLANES DE POLÍTICA FISCAL: EVALUACIÓN EX ANTE Y EX POST

Iván KATARYNIUK
Lucía LÓPEZ MUÑOZ
Javier J. PÉREZ

Banco de España

Resumen

El impacto y la eficacia de los planes presupuestarios, a nivel agregado, se encuentran íntimamente ligados a la transparencia en su definición, incluyendo la comunicación eficaz de los mismos y la credibilidad que los agentes económicos les asignen, a la fiabilidad de su implementación siguiendo las líneas aprobadas, así como a la rendición de cuentas, esto es, a la justificación de las desviaciones observadas *ex post* con respecto a los planes definidos *ex ante*, si estas se produjeran. En este artículo revisamos los elementos institucionales y técnicos que resultan más relevantes para la evaluación *ex ante* y *ex post* de los planes de política fiscal, desde una perspectiva macroeconómica.

Palabras clave: planes de política fiscal, presupuestos, desviaciones presupuestarias, evaluación de la política fiscal, previsiones macroeconómicas.

Abstract

The impact and effectiveness of budgetary plans, at the aggregate level, are closely linked to the transparency of their definition, including their effective communication and the credibility that economic agents assign to them, to the reliability of their implementation along the stated objectives, as well as to accountability, that is, to the justification of the deviations observed *ex post* with respect to the plans defined *ex ante*, should these occur. In this article we review the institutional and technical elements that are most relevant for the *ex-ante* and *ex-post* evaluation of fiscal policy plans from a macroeconomic perspective.

Keywords: fiscal policy plans, Government budget, budgetary deviations, fiscal policy evaluation, macroeconomic projections.

JEL classification: C53, E62, H60.

I. INTRODUCCIÓN

La rendición de cuentas de los gobernantes frente a los ciudadanos en el uso de los fondos públicos es una de las piezas fundamentales de la democracia, que en el caso de la política fiscal se refleja en el proceso de elaboración y ejecución de los presupuestos públicos dentro de los parámetros establecidos por la ley, y en un contexto de transparencia y control externo. La rendición de cuentas requiere una clara comprensión del impacto económico de la acción discrecional de los Gobiernos, manifestada en las políticas de ingresos y gastos públicos. Esto ha llevado gradualmente en las últimas décadas, y de manera más intensa en Europa desde mitad de la década de los años noventa del pasado siglo, al establecimiento de procedimientos e instituciones presupuestarias más sólidas, a la publicación de datos sobre la ejecución presupuestaria «cada vez más frecuentes» y sujetos a mejores estándares técnicos, y al desarrollo de herramientas cada vez más sofisticadas para realizar el seguimiento y la previsión de la actividad de las administraciones públicas (véase Leal *et al.*, 2008).

Las medidas de política fiscal se concretan, en condiciones normales, en el ciclo de elaboración

y aprobación del presupuesto anual que fija la legislación vigente, que puede incluir la definición de algunas actuaciones con compromisos de gasto en un horizonte plurianual (por ejemplo, en las áreas de infraestructuras o defensa). En situaciones extraordinarias, además, los Gobiernos pueden aprobar paquetes fiscales fuera de las fechas habituales, típicamente en respuesta a situaciones de crisis, como en marzo de 2020, en el contexto de la respuesta a la emergencia por la COVID-19; en diciembre de 2011, como reacción a las tensiones en los mercados de deuda soberana europea; o en noviembre de 2008, como respuesta a la crisis financiera global. Algunas actuaciones presentan, por su propia naturaleza, un carácter multianual, y afectan a varios ejercicios presupuestarios. Un ejemplo reciente de esto último son los proyectos a desarrollar en varios años mediante los fondos europeos aprobados bajo el paraguas del *Next Generation EU (NGEU)* (véase Alonso *et al.*, 2022).

El impacto y la eficacia de los planes presupuestarios, a nivel agregado, se encuentran íntimamente ligados a la transparencia en su definición, lo que viene condicionado, además, por la comunicación eficaz de los mismos y la credibilidad que los agentes económicos les asignen (véase End y Hong,

2021), a la fiabilidad de su implementación siguiendo las líneas aprobadas, así como a la rendición de cuentas, esto es, a la justificación de las desviaciones observadas *ex post* con respecto a los planes definidos *ex ante*, si estas se produjeran.

En las fases de anuncio y aprobación, los agentes tienen que evaluar los planes, es decir, examinarlos y asignar probabilidades de impacto a los mismos, bajo distintos supuestos sobre su definición y su ejecución, y tomar decisiones de consumo, inversión u oferta y demanda de trabajo, condicionadas a las sendas hipotéticas resultantes. En esta fase, en el caso español, cuentan con los informes correspondientes de la Autoridad Independiente de Responsabilidad Fiscal (AIReF), así como con la valoración que suelen proporcionar o publicar distintos analistas y organismos independientes (véase, por ejemplo, Hernández de Cos, 2021).

En la fase de implementación, los analistas tienen que evaluar, en tiempo real y con información incompleta y típicamente de publicación muy retrasada, el ajuste de las iniciativas anunciadas a actuaciones concretas, esto es, validar si lo que los Gobiernos anunciaron se corresponde con lo que los Gobiernos están haciendo, y actualizar las decisiones tomadas o adoptar otras nuevas (véase Paredes, Pérez y Pérez-Quirós, 2022).

Finalmente, *a posteriori*, una vez conocida la información necesaria, es posible evaluar el ajuste de los planes anunciados y aprobados a la implementación *ex post*, y en caso de que se hubieran producido desviaciones, analizar las razones de estas (véanse Leal *et al.*, 2008; Göttert y Lehmann, 2021). Por ejemplo, si estas se debieron a errores en las previsiones macroeconómicas y financieras en las que se basaron las estimaciones de ingresos públicos, a errores en la estimación de los multiplicadores fiscales asociados a las medidas, o a cambios discrecionales no anticipados en los planes de gasto y su composición (por ejemplo, ligados al ciclo electoral), entre otras posibles causas. En un proceso iterativo, el análisis de las desviaciones presupuestarias y sus motivos resulta crucial, más allá de su interés académico, para la credibilidad que los agentes económicos asignen a los nuevos anuncios de planes fiscales, lo que determinará su reacción ante estos.

En este artículo se revisa la experiencia disponible sobre estos temas. En primer lugar, se repasan el marco institucional europeo y nacional

para la elaboración y supervisión de los planes fiscales, incluyendo una referencia particular al papel de los organismos de control en los mismos (sección segunda). En segundo lugar, se exponen los elementos técnicos básicos necesarios, en nuestra opinión, para una adecuada evaluación *ex ante* de los planes (sección tercera). En tercer lugar se resaltan los aspectos más relevantes de la literatura de referencia sobre la evaluación *ex post* de los planes, elemento crucial para la forja de la credibilidad de los mismos y pieza esencial de la rendición de cuentas de los Gobiernos frente a los ciudadanos, destacando el caso del programa *NGEU* como un ejemplo prometededor de diseño institucional que incorpora incentivos y elementos de valoración *ex ante* y *ex post*. Finalmente, en la sección cuarta se presentan los principales mensajes del estudio.

II. LA ELABORACIÓN DE LOS PLANES DE POLÍTICA FISCAL EN EL MARCO DE LA UE

La literatura económica ha demostrado que la fortaleza del marco institucional es un factor de primer orden para explicar la calidad de los planes presupuestarios, tanto en lo referente a su diseño como a su cumplimiento (véanse, entre otros, Hallerberg, Strauch y Von Hagen, 2007; Reuter, 2019).

La legislación relevante y las instituciones que gobiernan los procesos presupuestarios nacionales en el caso español, y del resto de los países europeos, son nacionales. No obstante, presentan características comunes, derivadas de la necesaria coordinación de las políticas fiscales en el seno de la Unión Económica y Monetaria (UEM).

Según el marco que se estableció en el Tratado Constitutivo de la UEM, esta coordinación se limitaba a la publicación de los Programas de Estabilidad (o de convergencia, para los países de la UE que no habían adoptado la moneda única). Estos documentos, actualizados anualmente, presentan las medidas de política fiscal de los países y los objetivos presupuestarios a medio plazo a los que se comprometen para garantizar el cumplimiento de los parámetros establecidos en el Tratado. Estos se referían a la necesidad de evitar déficits excesivos, en particular, que el déficit público se situase en un valor inferior al 3 por 100 del PIB y dejaban a la Comisión Europea (CE) la supervisión de la evolución de la situación presupuestaria, así como del endeudamiento público, con el fin de detectar

errores manifiestos en la conducción de las finanzas públicas.

La crisis de la deuda soberana europea puso de manifiesto la necesidad de una coordinación más estrecha de las políticas económicas, dado que este marco fue insuficiente para garantizar la estabilidad presupuestaria, entre otros motivos, como consecuencia de políticas fiscales demasiado laxas en la época expansiva (véase García Perea y Gordo, 2016). Por ello, se introdujo una supervisión reforzada, organizada en torno al Semestre Europeo (1). Este mecanismo consta de dos fases diferenciadas. En el denominado «paquete de otoño», la CE emite su opinión referida a los borradores de plan presupuestario (o *DBP*, por sus siglas en inglés), en el que los Estados miembros (EE. MM.) presentan el escenario macroeconómico para el año siguiente, tanto del Gobierno central como de todos los subsectores de las administraciones públicas (AA. PP.), así como las principales medidas fiscales destinadas a cumplir los objetivos a medio plazo. Teniendo en cuenta esta opinión, los EE. MM. elaboran los Programas de Estabilidad, que deben fijar la senda nacional para cumplir el objetivo a medio plazo y ser coherentes con las recomendaciones relevantes emitidas por la CE y el Consejo. Además, los Programas de Estabilidad deben incluir indicaciones sobre el impacto esperado sobre las cuentas públicas, el crecimiento y el empleo de las reformas y medidas fiscales implementadas, así como los rendimientos económicos esperados de los proyectos de inversión pública que tengan un impacto presupuestario significativo.

En una segunda fase, la CE emite unas recomendaciones específicas de país, que deben aprobarse por el Consejo, en las que recomienda medidas que permitan aumentar la capacidad de crecimiento y empleo, así como mantener la solidez de las finanzas públicas. En el caso de que los objetivos planteados por un Estado miembro no se adecúen a las reglas fiscales derivadas del Tratado, la CE también puede abrir un procedimiento de déficit excesivo (PDE), por el cual el país se somete a una supervisión reforzada de sus finanzas públicas para devolverlas a una senda de sostenibilidad.

Junto a estas prácticas, las instituciones europeas expandieron las herramientas de seguimiento económico y presupuestario en dos dimensiones. Por una parte, ampliando los indicadores utilizados en el Semestre Europeo para tener en cuenta una visión más amplia, mediante el Procedimiento de desequilibrio macroeconómico (PDM). El PDM mo-

nitoriza la evolución de otros indicadores, además de los fiscales, como la deuda privada, la competitividad externa e interna o el desempleo, y examina en profundidad la situación de los EE. MM. con desequilibrios considerados excesivos. En caso de que el análisis de la CE concluya que se detecta un desequilibrio macroeconómico excesivo, el país debe tomar medidas para corregirlo, y, al igual que en el caso del PDE, exponerse a posibles sanciones si estas no son suficientes. Por otra parte, y como se comentará en las siguientes secciones, las reformas buscaron una mayor responsabilidad nacional en el diseño del marco fiscal, mediante la traslación del marco presupuestario europeo a la legislación nacional (el denominado *fiscal compact*) y la creación de instituciones fiscales independientes (parte de las reformas incluidas en el *two-pack*), con el mandato doble de supervisar el cumplimiento de las reglas fiscales numéricas y producir o avalar las proyecciones macroeconómicas en las que se basan los objetivos presupuestarios.

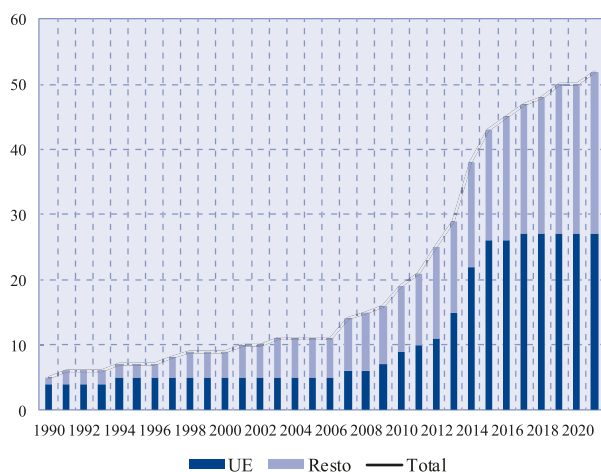
El encaje de esta legislación en el marco español se realizó en la Ley Orgánica 2/2012, de 27 de abril, de Estabilidad Presupuestaria y Sostenibilidad Financiera (LEP) (véase Hernández de Cos y Pérez, 2013).

El ámbito de los programas de estabilidad es el conjunto de las AA. PP., pero también recoge información sobre la contribución de sus diferentes subsectores. En el caso español, más allá de los objetivos presupuestarios para cada nivel de Gobierno, se proporcionan detalles sobre el comportamiento de las finanzas públicas de los niveles subcentrales, las comunidades autónomas y las corporaciones locales, así como información parcial sobre los planes de estos niveles de Gobierno. Esto distingue el Programa de Estabilidad español incluso de los de algunos de los países con un mayor nivel de descentralización fiscal en Europa, como es el caso de Alemania.

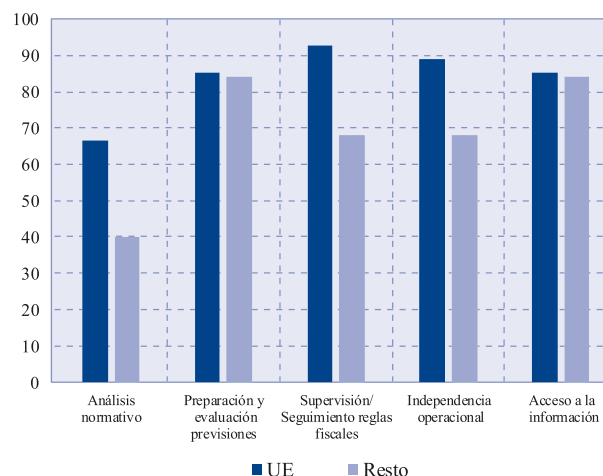
La importancia de los niveles subcentrales de Gobierno en el caso español se refleja en que, aunque la LEP asigna al Ministerio de Hacienda y Función Pública el papel de garantizar la aplicación de las normas, también juegan un papel destacado en el entramado institucional español los órganos colegiados de coordinación territorial —el «Consejo de Política Fiscal y Financiera de las comunidades autónomas» (CPFF) y la «Comisión Nacional de la Administración Local» (CNAL)—, así como los órganos interventores nacionales —en particular la Intervención General de la Administración del Estado (IGAE) y los propios

GRÁFICO 1
LAS INSTITUCIONES FISCALES INDEPENDIENTES

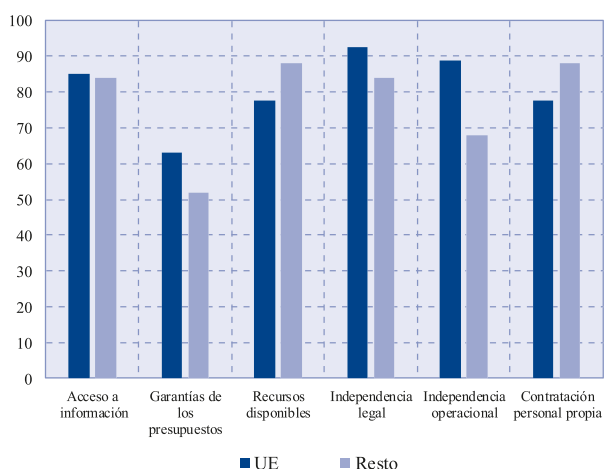
A. Aumento significativo en las últimas décadas



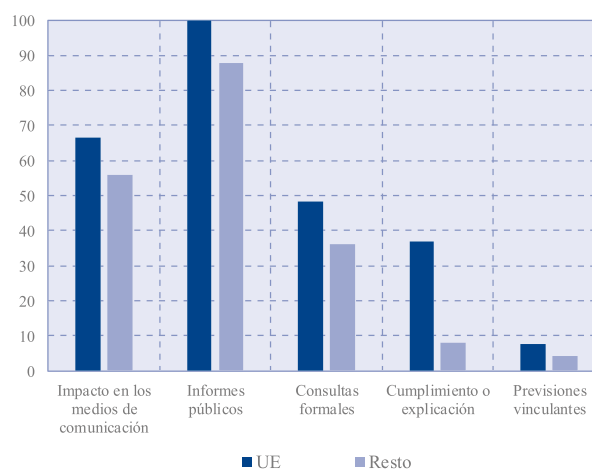
B. Porcentajes por tipo de actividad



C. Indicadores de independencia: porcentajes



D. Principales canales de influencia: porcentajes



Fuente: Fondo Monetario Internacional – «Fiscal councils dataset» (Davoodi *et al.*, 2022).

de cada región—, cámaras y sindicaturas de cuentas (véase Hernández de Cos y Pérez, 2014).

En este sentido, también tiene un papel destacado la AIREF. En efecto, más en general, en cuanto al control y validación de los planes presupuestarios reflejados en los programas de estabilidad, la reforma del marco de gobernanza europeo impulsó la creación de instituciones fiscales independientes (IFI) (véase Gordo, Hernández de Cos y Pérez, 2013), con el mandato explícito de evaluar las previsiones macroeconómicas que subyacen a estos, controlar el

grado de cumplimiento de las reglas fiscales numéricas nacionales y supranacionales, y llevar a cabo evaluaciones públicas sobre la necesidad de corregir una desviación significativa respecto del objetivo a medio plazo o de la trayectoria de ajuste para lograrlo.

Este modelo se aleja de las primeras IFI, cuyo mandato se centraba en la provisión de información y evaluación del impacto de las políticas públicas, como es el caso de Estados Unidos o Canadá, o en Europa, los Países Bajos. Así, las IFI europeas nacidas bajo el mandato del *two-pack* cubren una va-

riedad de funciones (véase AIReF, 2021; gráfico 1, paneles a y b): la elaboración de las previsiones macroeconómicas oficiales (Bélgica o Países Bajos), el análisis coste-beneficio o *spending reviews* (España o Países Bajos) o incluso la evaluación de programas electorales y acuerdos de coalición (Países Bajos, y, más recientemente, Letonia y Eslovenia). Como en el caso de los Programas de Estabilidad, en España la tarea de la AIReF cubre todos los niveles de las AA. PP., una práctica solo seguida, además, en Portugal y Eslovaquia.

En definitiva, se puede considerar que el objetivo de las reformas tras la crisis de la deuda soberana europea fue conseguir que los presupuestos nacionales, elaborados de forma anual, tuvieran en cuenta sus consecuencias intertemporales, reflejadas, fundamentalmente, en el nivel de deuda pública. Otra opción, implementada en algunos países, fue optar por presupuestos multianuales. Así, por ejemplo, en los Países Bajos, el presupuesto incluye un techo de gasto real plurianual, que se establece al comienzo de la legislatura, por un período igual o superior a esta. Fuera de Europa, otro ejemplo citado habitualmente es el caso de Uruguay, donde el Gobierno debe presentar un presupuesto para la legislatura completa, que es actualizado anualmente.

El éxito de estos marcos descansa, en todo caso, en su solidez y credibilidad institucional. Por ello, un importante aspecto de las IFI es su gobernanza interna, en aspectos como su independencia tanto funcional como económica (véase el gráfico 1, panel c), así como su estatus para poder cuestionar las previsiones del Gobierno. Un caso excepcional de este último procedimiento sucede en el Reino Unido, donde además de producir las previsiones macroeconómicas, la IFI también se encarga de elaborar las previsiones presupuestarias. Además, la CE ha recomendado la aplicación del principio de «cumplir o explicar» por el cual el Gobierno debe atender a las peticiones y requisitos emitidos por las IFI (véase el panel del gráfico 1). Sin embargo, la literatura señala que la aplicación de este principio a la elaboración de las previsiones ha sido heterogénea (AIReF, 2021). Además de esto, la literatura señala otros aspectos en los que las IFI pueden ejercer su labor de control, como la publicación de informes o a través de su impacto en los medios de comunicación (véase Davoodi *et al.*, 2022).

En todo caso, esta relación de supervisión entre IFI y Gobierno ha producido algunos episodios en

los que las previsiones fiscales de estos últimos no han sido avaladas. Por ejemplo, la IFI irlandesa (FCA) cuestionó las proyecciones gubernamentales de déficit y deuda en 2016, definidas como poco realistas, lo que desencadenó un debate en el Parlamento y la revisión de las proyecciones en los presupuestos de 2017. Asimismo, sucesos similares también se han producido entre el Gobierno italiano y la IFI de este país (PBO), en 2016, cuando las previsiones fueron validadas, pero con objeciones y, posteriormente, revisadas en el debate parlamentario. En 2018, la PBO rechazó las previsiones, lo que, seguido del dictamen negativo de la CE sobre el proyecto de plan presupuestario, obligó a un complicado proceso de revisión de las previsiones (véase Merlo y Fasone, 2021).

III. EVALUACIÓN EX ANTE DE LOS PLANES Y TRANSPARENCIA

Principios generales y modelos

La evaluación *ex ante* de los planes presupuestarios oficiales conlleva el análisis detallado de las siguientes piezas: el cuadro macroeconómico, la previsión de ingresos públicos, la coherencia entre el cuadro macroeconómico y la previsión de ingresos, y los elementos no discrecionales del gasto. Los elementos discrecionales del gasto público se toman como determinados en el análisis técnico de evaluación, en la medida en que forman parte del legítimo desarrollo de la política económica por parte del gobernante, aunque las limitaciones de los documentos públicos oficiales hacen necesarias, a veces, la elaboración de supuestos operativos (2).

Con respecto al cuadro macroeconómico, la valoración de las proyecciones oficiales debe partir de su comparación con las previsiones realizadas por otros analistas de manera independiente y con un historial de falta de sesgo –por ejemplo, el consenso de los analistas (3) o el Banco de España (4)–, o bien tomando como referencia un abanico amplio de modelos econométricos que permita evaluar la factibilidad de los objetivos a distintos horizontes (ver, por ejemplo, AIReF, 2018).

En relación con el resto de elementos resulta ilustrativo introducirlos haciendo uso de un modelo estándar de previsión de variables fiscales a horizonte un año del tipo (véase Leal *et al.*, 2008):

$$E_t[F_{t+1}] = \varepsilon E_t[M_{t+1}] + F_t^* \quad [1]$$

$$E_t[M_{t+1}] = m_1(L)M_t + m_2(L)F_t \quad [2]$$

Donde el operador $E_t[F_{t+1}]$ denota al valor esperado de una determinada variable a horizonte un año, L es el operador de retardos, y F hace referencia a las variables fiscales (de ingresos y gastos) que se determinan de acuerdo con una parte endógena, que depende de alguna (o algunas) variable macroeconómica de referencia, M , y un componente discrecional F^* . En un proceso típico de proyecciones, la solución del modelo formado por las ecuaciones [1] y [2] (estimado de manera conjunta para el conjunto del cuadro macroeconómico y todas las partidas relevantes de ingresos y gastos públicos) determinaría una previsión fiscal de referencia. El gráfico 2 describe, brevemente, los pasos seguidos. En primer lugar, las variables fiscales se determinan sobre la base de la evolución macroeconómica y de precios, multiplicadores fiscales, supuestos técnicos exógenos y medidas de política fiscal discrecionales. En segundo lugar, se generan los agregados fiscales y, por último, las variables fiscales sirven como *input* en los modelos macroeconómicos y de predicción.

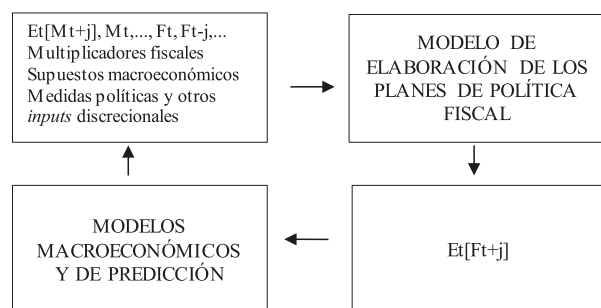
Con este marco de referencia a la vista, la previsión de una partida de los ingresos públicos, dado un cuadro macroeconómico, se determinaría por el producto de una elasticidad, ε , por su base imponible esperada de referencia, por ejemplo, las rentas laborales y del ahorro en el caso del IRPF, a lo que se sumarían, si los hubiera, aumentos adicionales derivados de medidas discrecionales de aumento de ingresos por cuantía F^* . Por tanto, el analista

debería validar la previsión de la base imponible presentada en el plan gubernamental, si la hubiera, de acuerdo con el cuadro macroeconómico de referencia, las cuantías proporcionadas debidas a los cambios legislativos, y que la elasticidad se encontrara dentro de los rangos estimados en la literatura económica de referencia. La detección de un sesgo al alza, por ejemplo, en la previsión oficial de una partida de ingresos podría provenir, por tanto, de una previsión optimista de las bases imponibles, del uso de unas elasticidades impositivas más altas que las representativas del consenso, de una sobreestimación de los ingresos derivados de los cambios normativos, o de una combinación de todos estos factores o alguno de ellos.

Con respecto a los elementos menos discrecionales del gasto, el procedimiento sería similar. En particular, la evolución del gasto en pensiones viene determinado esencialmente por la fórmula de revalorización establecida en la legislación y las variables demográficas que determinan la evolución de la población elegible; el subsidio de desempleo depende principalmente de la evolución del paro; y la carga de intereses refleja la evolución del endeudamiento público y de los tipos de interés.

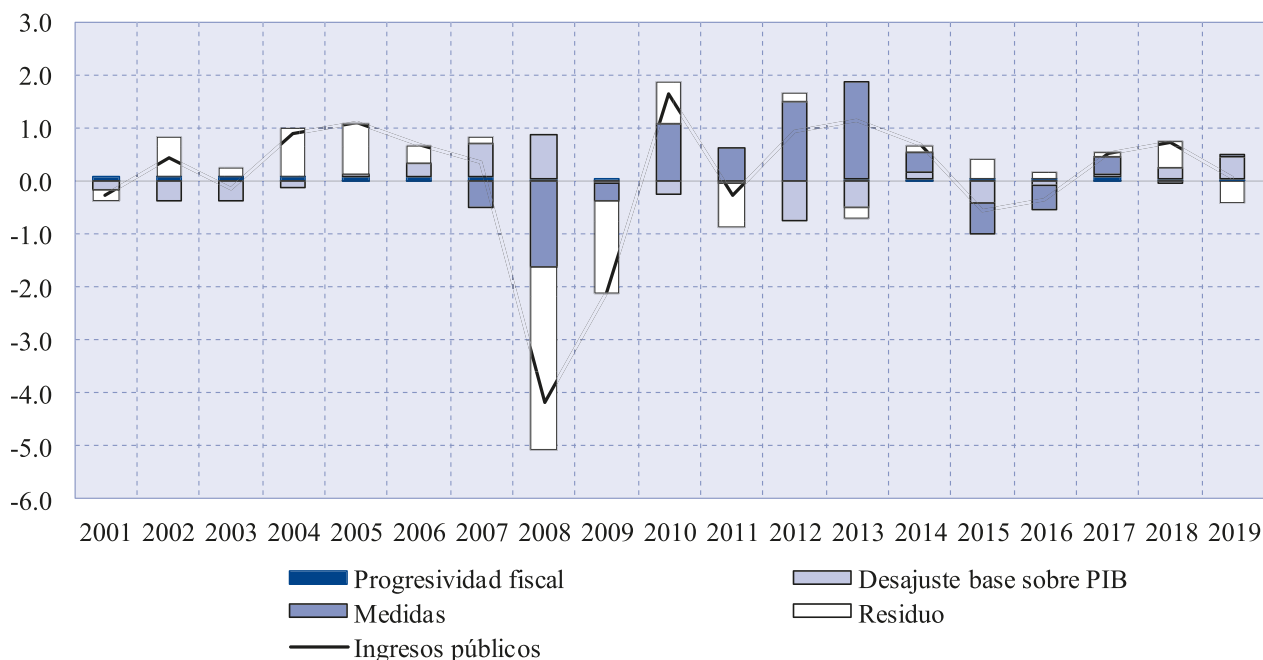
Volviendo a la valoración *ex ante* de las previsiones de ingresos públicos, en el gráfico 3 se presenta un modelo conceptual de descomposición de la ratio de ingresos públicos sobre el PIB, a lo largo del tiempo, en función de los principales factores macroeconómicos e institucionales que contribuyen a su variación. En concreto, estos factores se pueden resumir en (véase Burriel, López-Rodríguez y Pérez, 2017): 1) el desajuste en la evolución de las bases imponibles en relación al PIB nominal; 2) el efecto de la progresividad o *fiscal drag* asociada a impuestos cuya recaudación no evoluciona proporcionalmente a su base macroeconómica; 3) el impacto *ex ante* de medidas (cambios legislativos) de las reformas tributarias; y 4) un residuo agregado que recoge la variación no explicada de los factores anteriores. Así pues, en el gráfico 3 se muestra la variación de su ratio sobre el PIB del agregado de ingresos fiscales para la economía española durante el período 2001-2020. Esta descomposición también se puede usar para descomponer las previsiones tributarias oficiales y analizar su racionalidad. Por ejemplo, una valoración más reducida del impacto de los nuevos cambios legislativos sobre la recaudación, *ceteris paribus*, llevaría a un residuo (parte no explicada) muy elevado, lo que podría asociarse con un sesgo positivo en la estimación oficial de una determinada partida de ingresos.

GRÁFICO 2
PROCESO DE ELABORACIÓN DE LOS PLANES DE POLÍTICA FISCAL Y SU COHERENCIA CON LAS PREVISIONES MACROECONÓMICAS



Fuente: Elaboración propia basada en Leal *et al.* (2008).

GRÁFICO 3

**DESCOMPOSICIÓN DE LA VARIACIÓN DE LA RATIO DE INGRESOS SOBRE EL PIB
(Variación en porcentaje del PIB)**

Fuente: Elaboración propia basada en Burriel et al. (2017).

El caso del NGEU

Además de las herramientas derivadas de la literatura sobre predicción de variables fiscales que se han presentado en los párrafos anteriores, resulta ilustrativo presentar un nuevo procedimiento institucional de valoración recientemente diseñado en el marco del programa de fondos europeos *NGEU*.

En la sección segunda se ha destacado que las reformas del marco europeo han tendido a reforzar el control de la evaluación *ex ante* (a través de las opiniones realizadas por la CE sobre los planes presupuestarios o derivadas de la supervisión de las IFI nacionales). Sin embargo, la aprobación del Mecanismo de Recuperación y Resiliencia (MRR) del *NGEU* ha supuesto un importante cambio que ha afectado al diseño de las políticas fiscales y estructurales (véase Alonso y Matea, 2021) y ha desembocado en un proceso de evaluación *ex ante* de los objetivos de reforma de los países. Así, para acceder a los recursos del MRR, los EE. MM. deben detallar las inversiones y las reformas a las que des-

tinará los fondos en sus planes nacionales (Plan de Recuperación y Resiliencia, PRR). Estos PRR se han diseñado en un proceso de intenso diálogo con la CE, tanto en el período previo a la presentación de los PRR como en su evaluación, en el que la CE se ha asegurado de que los PRR cumplan las directrices generales del MRR.

Para orientar este diálogo, la CE publicó una serie de criterios con los que pretendía guiar el diseño de los PRR (5), en la medida en que utilizaría esos mismos criterios para certificar su aprobación y evaluar su cumplimiento. Así, además de los objetivos verdes y digitales del programa, la CE valoró, entre otros aspectos, la adecuación de las reformas a las recomendaciones específicas de país, así como la implementación del pilar europeo de derechos sociales. Asimismo, la liberación por parte de la CE de los desembolsos asociados a los PRR se encuentra vinculada al cumplimiento por parte del Estado miembro del marco de reglas fiscales. Así, en su artículo 10, el Reglamento señala que, en el caso de que el Consejo decida que un Estado miembro no ha tomado medidas efectivas para corregir su défi-

cit excesivo, la CE podrá presentar una propuesta al Consejo para suspender la totalidad o parte de los compromisos o pagos.

Junto a estos aspectos, de corte supranacional, los PRR recogen mecanismos de gobernanza y control específicos a nivel nacional. Entre otros, se refleja la necesidad de que las instituciones de auditoría nacionales o un comité creado para tal fin se encarguen del control interno de los planes. Adicionalmente, algunos países, como Italia y Portugal, han planteado estructuras de carácter horizontal, en el que grupos de interés y expertos externos participan en la implementación, y a veces en la evaluación, de los planes (véase Alonso *et al.*, 2022), mientras que en Grecia un organismo se dedica a la supervisión de las reformas y se prevé que estas sean certificadas por auditores independientes. En Italia, además, un organismo específico tiene el cometido de facilitar la eliminación de obstáculos administrativos para la implementación del plan.

En todo caso, la implementación de un programa del tamaño y la complejidad del *NGEU* requiere, en muchos casos, de la construcción de capacidad de gestión, tanto de las inversiones o reformas como de los datos. Por ello, varias reformas reflejadas en los PRR persiguen mejorar los sistemas de licitación pública o asegurar que se desplegará suficiente capacidad administrativa para garantizar un seguimiento adecuado (es el caso, por ejemplo, de Grecia).

Conscientes de este reto, las autoridades han diseñado, entre otros instrumentos, el *Recovery and Resilience Scoreboard*, para toda la información disponible sobre la implementación del *NGEU*. La información disponible, con carácter general, se puede agrupar en dos bloques. Por una parte, la información compilada por la CE sobre la implementación de los PRR, en base a los datos enviados por los EE. MM. sobre el cumplimiento de hitos y objetivos, y sobre los desembolsos realizados. La CE también ha desarrollado indicadores propios de gasto realizado, siguiendo una metodología transparente. Por otra parte, los EE. MM. tienen que proporcionar indicadores comunes sobre la implementación de los planes dos veces al año, en torno al 28 de febrero y en torno al 31 de agosto. La CE, en un siguiente paso, debe actualizar el *scoreboard* en abril y octubre, respectivamente.

IV. EVALUACIÓN EX POST DE LOS PLANES, RENDICIÓN DE CUENTAS Y CREDIBILIDAD

Lecciones de la literatura

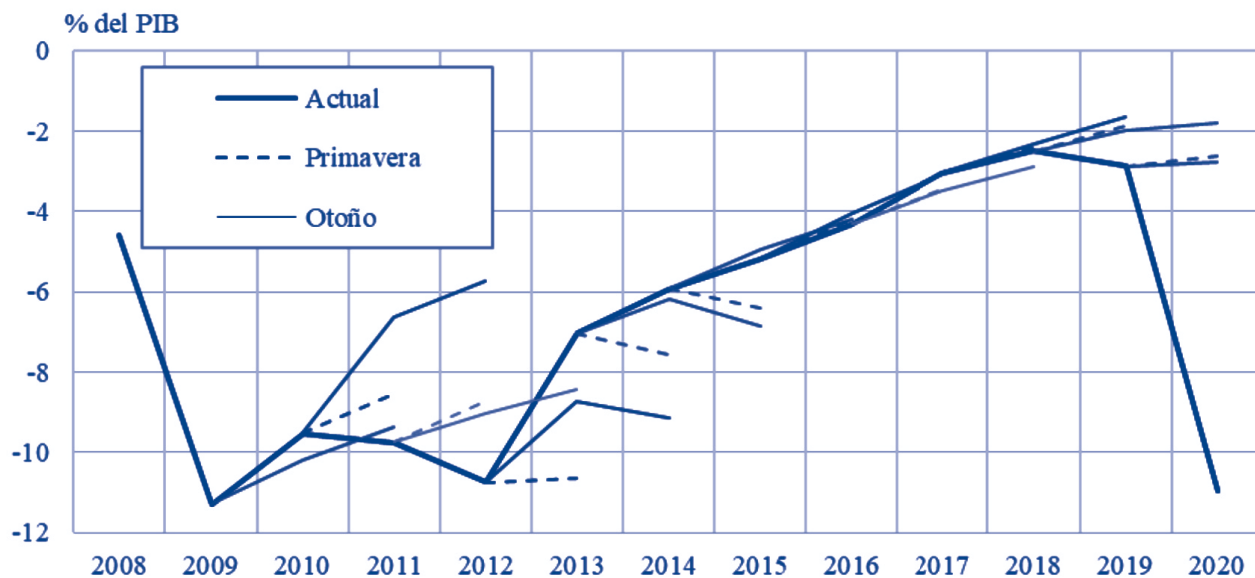
El seguimiento de los planes presupuestarios, entendido este como un ejercicio de valoración y evaluación en tiempo real de la adecuación de la ejecución de dichos presupuestos a los planes iniciales, resulta crucial para, entre otros aspectos, tratar de identificar a tiempo la existencia de riesgos de desviaciones de forma que se puedan articular las medidas necesarias durante el ejercicio para que estas no se produzcan. De hecho, una parte importante de la literatura más reciente se ha centrado en el desarrollo de modelos estadísticos en los que se utiliza como *input* los datos presupuestarios oficiales publicados con carácter mensual o trimestral, para evaluar la adecuación de esta información con los objetivos y planes anuales (véase Petropoulos *et al.*, 2022).

La incertidumbre sobre las políticas fiscales en tiempo real está íntimamente ligada a la credibilidad de los planes gubernamentales. Estos planes determinan cambios tributarios específicos y programas de gasto, así como episodios de consolidación fiscal, que configuran las decisiones y acciones de los agentes económicos. No obstante, la capacidad de los planes presupuestarios *ex ante* para transmitir información sobre el curso *ex post* de las políticas fiscales puede verse afectada por la presencia de sesgos políticos y comportamientos estratégicos de los Gobiernos, como lo muestra una rama bien establecida de la literatura (véase Paredes, Pérez y Pérez-Quirós, 2022, y las referencias que se proporcionan en este trabajo). De hecho, una gran parte de la literatura ha analizado desde un punto de vista teórico y empírico el sesgo potencial que el proceso institucional pueda tener sobre los planes de política fiscal del Gobierno y la naturaleza y propiedades de las desviaciones presupuestarias de los objetivos, así como de sus diferentes componentes. En la UE, la literatura empírica ha encontrado, de manera generalizada, que los ciclos políticos y una mayor debilidad de los marcos institucionales de reglas fiscales tienden a estar asociados a la aparición de sesgos al alza en las previsiones de ingresos públicos, esto es, a unas previsiones de ingresos que tienden a estar por encima de los ingresos que, *a posteriori*, acaban recibiendo, lo que se traduce en un sesgo a la baja en los déficits públicos y en los niveles de deuda pública.

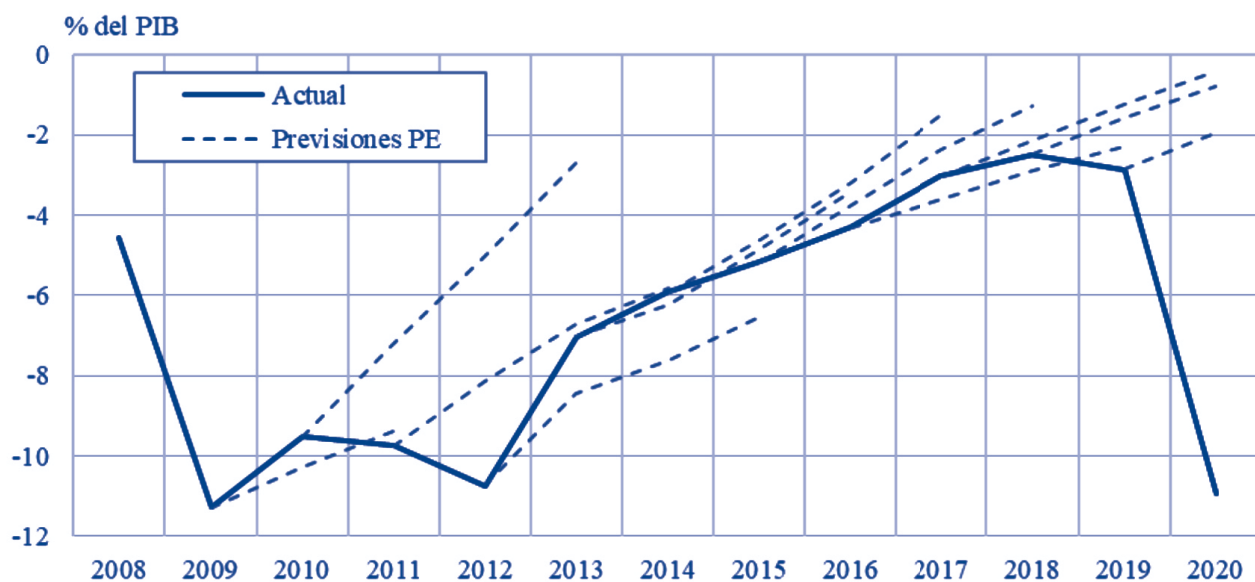
GRÁFICO 4

PREVISIONES DEL DÉFICIT DE LAS ADMINISTRACIONES PÚBLICAS ESPAÑOLAS

A. Comisión Europea



B. Gobierno de España (Programa de Estabilidad)



Nota: Las previsiones de la CE incluyen dos ejercicios para cada año «t» con el horizonte de proyección correspondiente. Asimismo, las series han sido ajustadas según los últimos datos publicados en el último ejercicio disponible de la CE.

Fuente: Previsiones de la CE - base de datos macroeconómica AMECO. Previsiones PE del Gobierno - proyecciones contenidas en el Programa de Estabilidad.

A modo ilustrativo, en el gráfico 4 se presenta el comportamiento a lo largo del tiempo de las previsiones y objetivos de déficit público de la CE y el Gobierno español para el período 2008-2020. En ambos paneles se observa cómo las previsiones parecen mostrar un cierto sesgo al compararlas con los datos observados, que depende de si la situación fiscal está mejorando o empeorando. En efecto, la literatura ha mostrado que en las situaciones en las que los déficits públicos empeoran, las previsiones suelen ser más optimistas, mientras que en épocas en las que mejoran, las previsiones suelen estar más ajustadas o incluso ser conservadoras. De manera crucial, las previsiones de la CE (representativas de las que realizan también otras organizaciones internacionales, como la OCDE o el FMI) tienden a mostrar un sesgo similar, aunque no tan marcado, como el del Gobierno. Numerosos estudios econométricos para paneles amplios de países de la UE muestran la presencia de estos sesgos, que indican la existencia de una correlación positiva entre «optimismo fiscal» y ciclos políticos, como se ha mencionado en el párrafo anterior. Lo que resulta más sorprendente es la presencia de este hecho en las previsiones de los organismos internacionales, lo que puede racionalizarse con una cierta dependencia de estas de algunos datos y estimaciones de medidas proporcionados por fuentes oficiales (véanse Merola y Pérez, 2013; Hannon, 2014).

Además de los sesgos institucionales, un factor fundamental para explicar las desviaciones presupuestarias son los errores en las previsiones macroeconómicas. Estos, en muchas ocasiones, se producen en situaciones de cambios de ciclo económico, o cuando se materializan perturbaciones macroeconómicas de gran tamaño, como la más reciente asociada al COVID-19, que ningún analista, público o privado, pudo anticipar. Esta situación se ilustra en el gráfico 5. En los paneles incluidos en esta figura se muestran diferentes previsiones de déficit público y crecimiento económico realizadas por la CE y el Gobierno español (Programas de Estabilidad) en tres episodios históricos. Por una parte, en el primer panel se muestra el período 2006-2008 en el que se observa una sorpresa positiva tanto en la ratio de déficit público como en crecimiento económico para los dos primeros años. Sin embargo, lo contrario sucede para 2008, donde tanto el Gobierno como la CE empeoraron sus últimas previsiones (otoño 2008) respecto a las anteriores (2006 y 2007) debido al reconocimiento del deterioro económico. Por otra parte, estos errores económicos debido a *shocks* inesperados se mostrarían de forma similar en los

dos paneles siguientes para los períodos 2009-2011 y 2018-2020, respectivamente.

La literatura reciente señala, además, las consecuencias de la falta de credibilidad de los planes fiscales. Así, End y Hong (2021) señalan que los Gobiernos, en general, prevén déficits fiscales inferiores a los previstos por el sector privado, y que los errores de previsión del Gobierno impactan sobre las previsiones futuras del sector privado, que se vuelven más pesimistas. Este estudio calcula que esta falta de credibilidad tiene además un coste financiero: por cada 2 puntos del PIB de discrepancia en la previsión del déficit fiscal, se produce un aumento de 6 puntos básicos del coste de la deuda pública (FMI, 2021).

Un aspecto adicional de particular relevancia en la Unión Europea es la necesidad de disponer de estadísticas fiables e independientes tanto para las variables fiscales como para las incluidas en el PDM, de manera que incluso el *six-pack* incorpora un régimen de sanciones por el incumplimiento de los requerimientos estadísticos (6). Estas reformas vinieron motivadas, principalmente, por los problemas con las estadísticas de déficit público de Grecia que se evidenciaron en la época de su acceso al área del euro y con posterioridad (7). El gráfico 6 muestra las diversas revisiones del saldo público en Grecia en porcentaje del PIB en una muestra temporal muy amplia. Grecia se enfrentó a un problema de credibilidad debido a los errores descubiertos en sus estadísticas y sus revisiones frecuentes hasta 2010-2011, cuando la CE reconoció públicamente el problema que ello suponía para la credibilidad de todo el sistema estadístico europeo. De hecho, es en la segunda década dónde se observa un mayor ajuste de los datos, reduciéndose la dispersión de las diferentes revisiones para cada año. Esto se constata en que la desviación promedio de las previsiones en la primera década analizada es cinco veces superior a la de la segunda. No obstante, en el período 2000-2008, la existencia de revisiones estratégicas de los datos presupuestarios no se ciñó solamente al caso de Grecia, sino que tuvo un carácter general (véase de Castro, Rodríguez-Vives y Pérez, 2013).

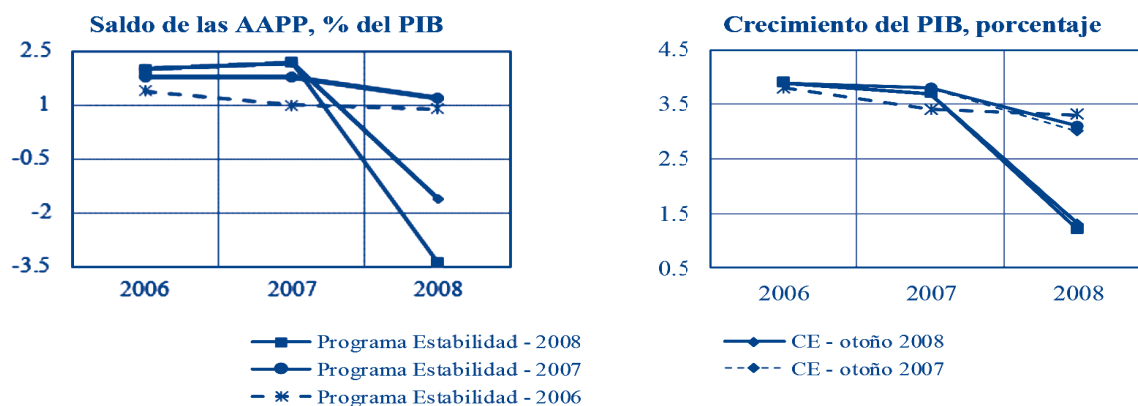
El caso del NGEU

La experiencia con el MRR debería permitir, además, avanzar en el diseño de planes creíbles. En primer lugar, el MRR supone un avance en la coordinación de las políticas a nivel supranacional, con la CE ejerciendo un rol centralizador de las diferentes

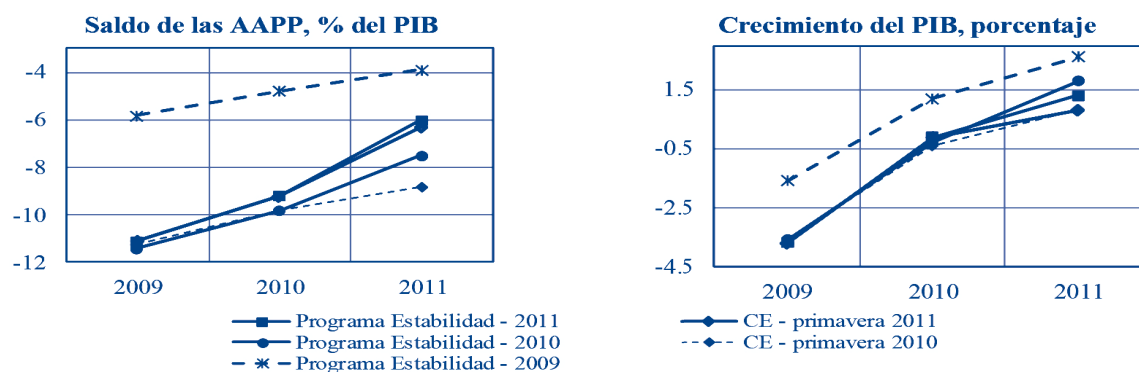
GRÁFICO 5

PREVISIONES DE DÉFICIT DE LAS AA. PP. Y DE CRECIMIENTO DEL PIB PUBLICADAS EN EL MOMENTO PREVIO A VARIOS EPISODIOS DE CRISIS RECIENTES

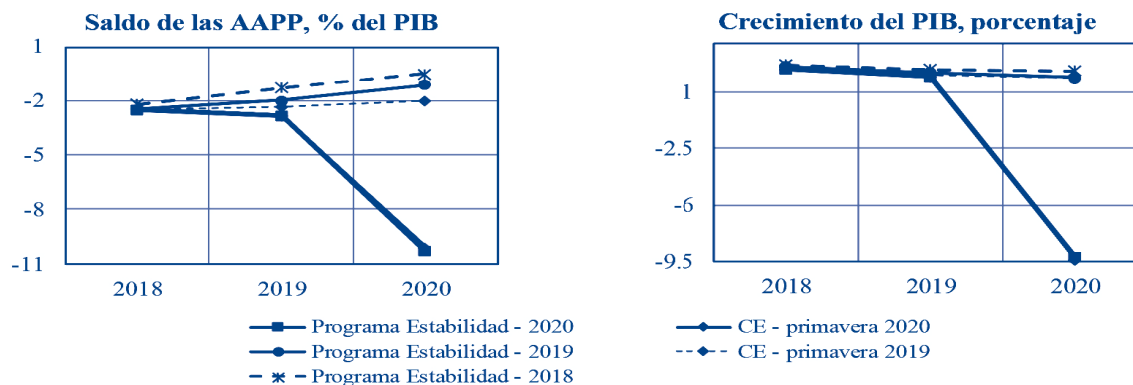
A. La crisis financiera global 2008



B. La crisis de deuda soberana del área del euro



C. La crisis del COVID-19

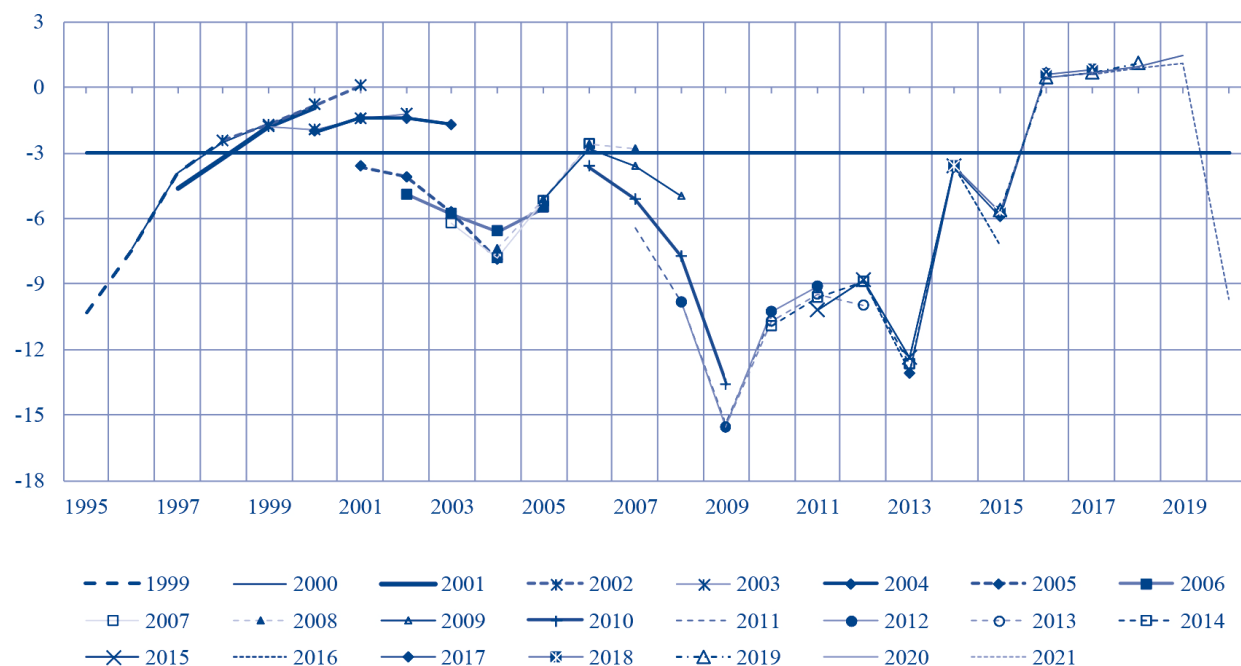


Nota: Las previsiones de la CE incluyen dos ejercicios para cada año «t» con el horizonte de proyección correspondiente.

Fuente: Previsiones de la CE - base de datos macroeconómica AMECO. Previsiones PE del Gobierno - proyecciones contenidas en el Programa de Estabilidad.

GRÁFICO 6

LAS REVISIONES DE LOS DATOS DE DÉFICIT PÚBLICO: EL CASO DE GRECIA (Porcentaje del PIB, datos publicados en primavera de cada año)



Fuente: Elaboración propia basada en De Castro *et al.* (2013) y actualización de datos proporcionada por estos autores.

inversiones y reformas realizadas por los EE. MM. En segundo lugar, la existencia de incentivos monetarios para el cumplimiento del calendario acordado puede ayudar a sortear los tradicionales obstáculos internos a la realización de reformas, lo que permitiría un mejor ajuste entre los objetivos planteados y la realización *ex post*. Finalmente, el calendario de hitos y objetivos permite un seguimiento cercano de los avances y la posibilidad de que el no cumplimiento con las directrices acordadas *ex ante* tenga consecuencias (por ejemplo, mediante el retraso en la aprobación de nuevos fondos, una circunstancia que ya se ha producido en los casos de los PRR de Hungría y Polonia).

El MRR, de esta forma, acerca a la Unión Europea a ejercer un rol de presupuesto federal, en el que se financien programas a gran escala que sean de interés común para los EE. MM. Este rol es el que ha adoptado el Gobierno federal en Estados Unidos desde el siglo XIX para impulsar capítulos de gasto que estuvieran sujetos a un estándar de calidad común (véase Maxwell, 1952). De esta forma, se reconoce la necesidad de utilizar los recursos en común para im-

pulsar programas de interés para todos los miembros (en particular, aquellos con importantes externalidades hacia el resto de socios, como la lucha contra el cambio climático), y las subvenciones funcionan como un vehículo intermedio que permite a los EE. MM. continuar administrando esas funciones, a la vez que se reconoce la autoridad federal para imponer condiciones y evaluar su cumplimiento.

El éxito y la extensión de esta función dependerá, en cualquier caso, tanto del cumplimiento de esos estándares de calidad (esto es, de la aprobación y la profundidad de las inversiones y las reformas contenidas en los PRR) como de la capacidad en el largo plazo de la UE de ampliar sus fuentes de ingresos, lo que permitiría dar continuidad a esta forma de colaboración entre niveles de Gobierno.

V. PRINCIPALES MENSAJES

En este artículo hemos argumentado que el impacto y la eficacia de los planes presupuestarios, a nivel agregado, se encuentran íntimamente ligados

a la transparencia en su definición, incluyendo la comunicación eficaz de los mismos y la credibilidad que los agentes económicos les asignen, a la fiabilidad de su implementación siguiendo las líneas aprobadas, así como a la rendición de cuentas, esto es, a la justificación de las desviaciones observadas *ex post* con respecto a los planes definidos *ex ante*, si estas se produjeran. Todos estos elementos se encuentran condicionados por la calidad del marco institucional en el que se desenvuelven y la calidad de los datos disponibles para la evaluación. Desde el punto de vista macroeconómico (agregado), la evaluación *ex ante* de los planes requiere un análisis detallado de numerosos elementos técnicos que permitan, en particular, evaluar la coherencia de los planes con las previsiones macroeconómicas que los acompañan, mientras que la evaluación *ex post* requiere un análisis exhaustivo de las razones que justifican y explican las desviaciones presupuestarias.

Desde esta perspectiva, en el contexto europeo, la puesta en marcha del *NGEU* supone un refuerzo del marco de supervisión vigente, en tanto en cuanto una evaluación favorable *ex ante* de los planes ha sido una condición necesaria para la liberación inicial de los fondos, mientras que la continuidad de estos, se condiciona a una evaluación *ex post* favorable, basada en el cumplimiento de los compromisos adquiridos inicialmente. Este sistema puede servir como inspiración para algunos elementos de la necesaria reforma del marco europeo de gobernanza económica, que, a fecha de cierre de este artículo, se encuentra en discusión (véase Alloza *et al.*, 2021).

NOTAS

(1) Reglamento (UE) n.º 473/2013 del Parlamento Europeo y del Consejo de 21 de mayo de 2013 sobre disposiciones comunes para el seguimiento y la evaluación de los proyectos de planes presupuestarios y para la corrección del déficit excesivo de los Estados miembros de la zona del euro.

(2) Por esto queremos decir que, en algunas ocasiones, tomando como ejemplo el caso español, los planes de gasto de muchas partidas discrecionales se fijan en términos de contabilidad presupuestaria, sin un paralelo directo con los grandes agregados de contabilidad nacional, que es el estándar con el que se definen los objetivos de estabilidad. Por ejemplo, típicamente, los Presupuestos Generales del Estado presentan unos planes de inversión pública en términos de caja para el subsector Estado, que se agregan con los planes del sector corporativo estatal (parte del cual no se consolida dentro del sector de las AA. PP.), mientras que las comunidades autónomas y las corporaciones locales hacen algo similar en sus presupuestos. La conexión entre esta información y el plan agregado en contabilidad nacional para la formación bruta de capital fijo de las AA. PP. que se presenta resulta compleja para los analistas externos a la elaboración de los presupuestos.

(3) Por ejemplo, las realizadas por el panel de Funcas, que se publican en el sitio web: <https://www.funcas.es/coyuntura-economica/espana/panel-de-previsiones/panel-de-previsiones/>.

(4) En este caso, las previsiones que se publican de manera trimestral se pueden consultar en el espacio web: <https://www.bde.es/bde/es/secciones/informes/analisis-economico-e-investigacion/proyecciones-macro/relacionados/boletin-economico/informes-de-proyecciones-de-la-economia-espanola/>.

(5) Véase Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021 por el que se establece el Mecanismo de Recuperación y Resiliencia.

(6) Véase Reglamento (UE) n.º 1175/2011 del Parlamento Europeo y del Consejo de 16 de noviembre de 2011.

(7) Véase European Commission. Brussels, 8.1.2010. COM(2010) 1 final. Report on Greek Government Deficit and Debt Statistics.

BIBLIOGRAFÍA

- AlReF (2021). Las Instituciones Fiscales Independientes en la Unión Europea. *Documento Técnico*, 1/2021.
- AlReF (2018). *Metodología. Escenario macroeconómico y presupuestario 2018 y 2019 bajo el supuesto de políticas constantes*. Centro Documental de la AlReF.
- ALLOZA, M., ANDRÉS, J., BURRIEL, P., KATARYNIUK, I., PÉREZ, J. J. y VEGA, J. L. (2021). The reform of the European Union's fiscal governance framework in a new macroeconomic environment. *Bank of Spain Occasional Paper Series*, 2121.
- ALONSO, D. y MATEA, M. L. (2021). El semestre europeo 2021 y el plan de recuperación, transformación y resiliencia. *Boletín Económico*, 4/2021, Banco de España.
- ALONSO, D., KATARYNIUK, I., MORENO, C. y PÉREZ, J. J. (2022). El programa *Next Generation EU*: características y claves para su éxito. *Información Comercial Española*. Próxima publicación.
- BURRIEL, P., LÓPEZ, D. y PÉREZ, J. J. (2017). Evaluación macroeconómica de las reformas impositivas: aspectos metodológicos y algunas aplicaciones. *Papeles de Economía Española*, 154.
- DAVOODI, H., ELGER, P., FOTIOU, A., GARCIA-MACIA, D., LAGERBORG, A., LAM, R. y PILLAI, S. (2022). *Fiscal Councils Dataset: The 2021 Update*. Washington D. C.: International Monetary Fund.
- CASTRO, F. DE, PÉREZ, J. J. y RODRÍGUEZ-VIVES, M. (2013). Fiscal data revisions in Europe. *Journal of Money, Credit and Banking*, 45, pp. 1189-1211.
- END, N. y HONG, G. H. (2021). *Trust What You Hear: Policy Communication and Fiscal Credibility*. Fondo Monetario Internacional. Mimeo.
- GORDO, E., HERNÁNDEZ DE COS, P. y PÉREZ, J. J. (2015). Instituciones fiscales independientes en España y en la UEM. *Boletín Económico del Banco de España*, febrero, pp. 27-44.
- GÖTTERT, M. y LEHMANN, R. (2021). Tax Revenue Forecast Errors: Wrong Predictions of the Tax Base or the Elasticity? *CESifo Working Paper*, n.º 9148.
- HALLERBERG, M., STRAUCH, R. y VON HAGEN, J. (2007). The design of fiscal rules and forms of governance in European Union countries. *European Journal of Political Economy*, 23, pp. 338-359.

HANNON, P. (2014). Difficulties in Forecasting Borrowing Requirements. *The Wall Street Journal*, 10 de enero, <https://www.wsj.com/articles/BL-MBB-14492>

HERNÁNDEZ DE COS, P. (2021). Comparecencias ante la Comisión de Presupuestos del Congreso de los Diputados, el 25 de octubre de 2021, y ante la Comisión de Presupuestos del Senado, el 30 de noviembre de 2021, en relación con el Proyecto de Presupuestos Generales del Estado para 2022. *Documentos Ocasionales del Banco de España*, 2130.

HERNÁNDEZ DE COS, P. y PÉREZ, J. J. (2013). La nueva Ley de Estabilidad Presupuestaria. *Boletín Económico*, 4/2013, pp. 65-78. Banco de España.

HERNÁNDEZ DE COS, P. y PÉREZ, J. J. (2014). El seguimiento en tiempo real de los presupuestos autonómicos y locales. *Presupuesto y Gasto Público*, 77, pp. 79-99.

LEAL, T., PÉREZ, J. J., TUJULA, M. y VIDAL, J.-P. (2008). Fiscal forecasting: lessons from the literature and challenges. *Fiscal Studies*, 29, pp. 347-386.

MAXWELL, J. A. (1952). Brief history of grants. En *Federal Grants and the Business Cycle*. National Bureau of Economic Research (NBER).

MERLO, S. y FASONE, C. (2021). Differentiated Fiscal Surveillance and the Democratic Promise of Independent Fiscal Institutions in the Economic and Monetary Union. *Swiss Political Science Review*, 27, pp. 582-600.

MEROLA, R. y PÉREZ, J. J. (2013). Fiscal forecast errors: governments versus independent agencies? *European Journal of Political Economy*, 32, pp. 285-299.

PAREDES, J., PÉREZ, J. J. y PÉREZ-QUIRÓS, G. (2022). *Fiscal targets. A guide to forecasters?* Banco de España.

PETROPOULOS, F. et al. (2022). Forecasting in practice. *International Journal of Forecasting*. En prensa.

REUTER, W. M. (2019). When and why do countries break their national fiscal rules? *European Journal of Political Economy*, 57, pp. 125-141.

COLABORACIONES

III.
GASTO PÚBLICO

Resumen

Este artículo analiza en primer lugar por qué es importante evaluar el funcionamiento del sistema judicial. Así, recuerda que, además de ser esencial desde un punto de vista jurídico, la justicia es un determinante de la productividad de la economía. A continuación, esta investigación expone cuáles son las mejores medidas de eficiencia judicial y plantea cómo cuantificarlas en la práctica, con base en estadísticas nacionales e internacionales. Finalmente, tras destacar algunos resultados cuantitativos para España, este papel repasa dos aplicaciones prácticas: el impacto de la complejidad de la regulación en la justicia y cómo la ineficacia de la justicia tiene implicaciones negativas en la inversión a nivel de empresa en España.

Palabras clave: evaluación de políticas públicas, administración de justicia, sistema judicial, impactos económicos, España.

Abstract

This article first analyzes why it is important to evaluate the functioning of the judicial system. It recalls that, in addition to being essential from a legal standpoint, judicial efficacy is a determinant of an economy's productivity. This research then sets out the best measures of judicial efficacy and proposes how to quantify them in practice, based on national and international statistics. Finally, after highlighting some quantitative results for Spain, this paper reviews two practical applications: the impact of regulatory complexity on justice and how the inefficacy of justice has implications for firm-level investment in Spain.

Keywords: public policy evaluation, justice, judicial system, economic impacts, Spain.

JEL classification: D02, D04, K40, O10.

LA EVALUACIÓN DE LAS POLÍTICAS PÚBLICAS: LA ADMINISTRACIÓN DE JUSTICIA

Juan S. MORA-SANGUINETTI (*)

Banco de España e ICAM

I. ¿POR QUÉ ES NECESARIO EVALUAR EL FUNCIONAMIENTO DE LA ADMINISTRACIÓN DE JUSTICIA?

DESDE un punto de vista jurídico, el sistema judicial es un instrumento fundamental de garantía de derechos y de libertades en una sociedad desarrollada. El uso de la justicia (la litigación) es, de hecho, un derecho fundamental, reconocido en la Constitución española (derecho a obtener la tutela efectiva de los jueces y tribunales) y por el Convenio Europeo de Derechos Humanos (el denominado «derecho a un proceso equitativo») (1).

Cabe hablar de dos situaciones generales de «indefensión» (utilizando el término recogido por la Constitución española) a las que se puede enfrentar un ciudadano o una empresa y para las que puede necesitar de la asistencia del sistema judicial: la que se puede producir cuando los ciudadanos o las empresas actúan en un ámbito privado y se enfrentan a las acciones de otros sujetos (privados). Por otro lado, esos mismos ciudadanos o empresas pueden verse enfrentados a la actuación incorrecta por parte de los poderes públicos.

En el primero de esos dos casos, cabe pensar de una multitud de circunstancias con una importante repercusión econó-

mica: desde la poca entidad que pueda tener la compra de una barra de pan hasta contratos mucho más complejos como la adquisición de maquinaria para la industria o la compraventa de un apartamento. El sistema judicial actúa como un instrumento de disciplina y mejora de la contratación privada.

Por el lado público, el sistema judicial defiende a los ciudadanos y empresas de los riesgos de «expropiación» o favoritismo hacia determinados intereses por parte del Gobierno. Los competidores que resultan desfavorecidos se ven desposeídos de parte de su capacidad para aportar valor al mercado (véase Mora-Sanguinetti, 2022).

Por todo ello, desde esa perspectiva general o jurídica, es esencial que el sistema judicial funcione correctamente (de forma eficaz) y, por tanto, es crítico que se realice una adecuada evaluación de su eficacia. Sin embargo, es importante recordar que esas dos perspectivas (la privada y la pública) tienen una alta relevancia desde el punto de vista del análisis económico y ya fueron puestas en valor por Coase (1960) y Acemoglu y Johnson (2005).

Por esos motivos, pero centrandó la discusión en el análisis económico, es posible señalar que el funcionamiento de la justicia determina parcialmente el

CUADRO N.º 1

ESTUDIOS SOBRE LOS IMPACTOS ECONÓMICOS DEL FUNCIONAMIENTO DE LA JUSTICIA

MATERIA	EJEMPLOS DE ESTUDIOS ESPECÍFICOS A ESCALA INTERNACIONAL	EJEMPLOS DE ESTUDIOS ESPECÍFICOS PARA ESPAÑA
Mercados de crédito	Fabbri y Padula (2004); Jappelli, Pagano y Bianco (2005); Shvets (2013)	Fabbri (2010); Mora-Sanguinetti, Martínez-Matute y García-Posada (2017)
Crecimiento, tamaño y especialización empresarial	Beck, Demirguc-Kunt y Maksimovic (2006); Laeven y Woodruff (2007); Giacomelli y Menon (2017)	García-Posada y Mora-Sanguinetti (2015a)
Inversión y especialización	Nunn (2007); Chemin (2012)	Dejuán y Mora-Sanguinetti (2021)
Distorsión de relaciones laborales	Berger y Neugart (2011); Marinescu (2011)	Jimeno, Martínez-Matute y Mora-Sanguinetti (2020)
Mercado de vivienda	Casas-Arce y Saiz (2010)	Mora-Sanguinetti (2012)
Entrada de empresas y emprendimiento	Desai, Gompers y Lerner (2005); Stephen, Urbano y van Hemmen (2009); Ardagna y Lusardi (2010)	García-Posada y Mora-Sanguinetti (2015b)

Fuentes: Elaboración propia, Mora-Sanguinetti (2020) y 2022).

comportamiento de la productividad, es decir, influye en que una economía pueda crear riqueza en el largo plazo. Entre otras razones esto ocurre porque el buen funcionamiento de la justicia facilita el crecimiento de las empresas (García-Posada y Mora-Sanguinetti, 2015a). A su vez, las empresas de un tamaño superior muestran un mejor rendimiento de las inversiones en I+D. Por otro lado, una justicia eficaz también se relaciona con la salida de las empresas ineficientes del mercado. La calidad de la justicia se relaciona, en fin, con una mayor seguridad jurídica, lo que sostiene la inversión (Dejuán y Mora-Sanguinetti, 2021), una más fácil combinación de factores productivos y la toma de decisiones empresariales.

El conjunto de impactos o de relaciones entre el funcionamiento de la justicia y los distintos mercados o facetas de la economía es incluso más amplio, pudiéndose encontrar también el efecto en la disponibilidad de crédito (Mora-Sanguinetti, Martínez-Matute y García-Posada, 2017).

La bibliografía mencionada en estos ejemplos hace mención a España, pero esos resultados son coherentes con otros obtenidos por otros estudios a nivel internacional (por ejemplo, Jappelli, Pagano y Bianco [2005] o Laeven y Woodruff [2007]).

A modo de resumen de diferentes estudios que analizan los impactos de los sistemas judiciales en distintos aspectos del funcionamiento de una economía puede consultarse el cuadro n.º 1. En primer lugar, se mencionan los trabajos realizados para otras economías y, a continuación, algunos resultados para España.

Vistos los resultados anteriores, fundamentados con distintos estudios publicados, no debería resultar sorprendente la reflexión de Douglass C. North (véase, entre otras referencias, North [1990]), premio Nobel en Economía de 1993, según la cual defendió que el factor más importante de desarrollo económico de largo plazo es el buen funcionamiento de las «instituciones de ejecución». Es

decir, esa sería la principal diferencia en el marco institucional de las economías desarrolladas comparadas con aquellas que conforman el «tercer mundo». Esas «instituciones de ejecución» incluyen todos los mecanismos dedicados a hacer cumplir la ley y los contratos, siendo su ejemplo representativo el sistema judicial.

II. LA EVALUACIÓN DE LA JUSTICIA A ESCALA INTERNACIONAL Y SU METODOLOGÍA

Como se introducía, los impactos económicos del sistema judicial (además de los propiamente jurídicos) justifican la medición y evaluación de su calidad y de su funcionamiento.

En general, la bibliografía a nivel internacional reduce la evaluación a tres parámetros: la medición de la lentitud de su funcionamiento, su coste o su predictibilidad. Como veremos, esas tres medidas a su vez suelen ceder ante la primera. La lentitud parece, pues, ser el indicador

global más adecuado de calidad o eficacia judicial.

1. Fuentes de información sobre la eficacia de la justicia a nivel internacional

Cabe destacar tres fuentes de información estadística a nivel internacional, todas ellas respaldadas por sus respectivas metodologías: las proporcionadas por la Comisión Europea para la Eficacia de la Justicia (CEPEJ), cuyos últimos informes completos son de 2016 y 2018, pero ha publicado alguna información en 2020, el *OECD Civil Justice Project* elaborado por la OCDE en 2013 (Palumbo *et al.*, 2013a y 2013b) en la que participé como coautor y, en tercer lugar, el proyecto *Doing Business (DB)* del Banco Mundial (indicador de «cumplimiento de contratos»/«enforcing contracts»). Su último informe es de 2020.

Las dos primeras fuentes (CEPEJ y OCDE) comparten mayormente la misma metodología y se basan en información real (medible) de los sistemas judiciales de los países que cubren. En el primer caso de los países europeos, en el segundo también se analizan algunos países de la OCDE fuera de Europa. La tercera fuente (DB) se fundamenta parcialmente en percepciones, su metodología estuvo basada originalmente en el trabajo de Djankov *et al.* (2003) y cubre una gran cantidad de países, europeos y no europeos.

El DB es, tal vez, la fuente de información más utilizada e influyente, pero puede contener sesgos desfavorables a sistemas jurídicos de familia romana-francesa, como el español (2). El 16 de septiembre de 2021 el Banco

Mundial comunicó que el DB va a dejar de elaborarse.

2. Metodología

Existe una multiplicidad de perspectivas desde las cuales podría evaluarse el buen o mal funcionamiento de un sistema judicial, incluyendo su medición con datos objetivos (por ejemplo, la rapidez) o bien su estudio mediante la captación de percepciones (por ejemplo, de ciudadanos y empresas), que podrían no coincidir con esas primeras mediciones (percepción de injusticia o de falta de independencia) (3).

Desde un punto de vista objetivo, los distintos estudios suelen limitar el análisis a tres elementos: la rapidez (lentitud), el coste y la predictibilidad de la justicia. El primero hace referencia al tiempo necesario para que un sistema judicial emita una resolución para un caso (conflicto) concreto, el segundo al dinero o recursos que las partes tienen que gastar (invertir) para alcanzar esa resolución y el tercero a si la resolución es o no esperable. Esa tercera perspectiva es la más difícil de medir. Estas tres mediciones pueden ser relacionadas de forma sencilla con los impactos económicos de un sistema judicial.

Más en concreto, Palumbo *et al.* (2013a), en el contexto del *OECD Civil Justice Project*, midieron la lentitud como el número de días necesarios para superar una (o más instancias) en un procedimiento dentro de la jurisdicción civil (*trial length*). Para la medida de coste, tomaron información de los indicadores DB, proporcionando una aproximación al coste privado total descontado por el valor espera-

do de la asistencia jurídica. Para la predictibilidad, se midieron el número de recursos que se producen de las decisiones judiciales en instancias superiores. La OCDE argumenta que hay más recursos si las partes tienen dudas sobre el sentido de la resolución judicial en la instancia superior.

Lograr mejoras en las tres dimensiones (lentitud, coste y predictibilidad) es difícil pues son perspectivas que pueden presentarse enfrentadas: un sistema judicial cuyo uso implica un bajo coste puede atraer una gran cantidad de casos (conflictos) a ser resueltos en él, provocando un aumento de la congestión y un empeoramiento en la primera dimensión (4).

En España concretamente, otros informes han puesto también atención a estas dimensiones. El Círculo de Empresarios (2003) apuntaba a que el sistema judicial español tenía un nivel medio o bajo de satisfacción para las empresas que eran miembros del Círculo basándose precisamente en medidas de lentitud y predictibilidad de las sentencias. Para más detalles, puede consultarse Mora-Sanguinetti (2010b).

Diferentes análisis se concentran, sin embargo, en una medida única de «calidad» o eficacia del sistema judicial, que es la lentitud pues concentra información relacionada con las tres dimensiones. Un sistema judicial muy lento también es muy costoso (requiere una inversión de recursos para las partes dilatada en el tiempo, por ejemplo, el pago de los abogados) y es, por definición, impredecible. El análisis de los datos pone de manifiesto, de hecho, que el coste de

un procedimiento judicial (una vez tenida en cuenta la posible existencia de asistencia jurídica gratuita) y su duración están altamente relacionados. La correlación se encuentra entre 0,31 y un 0,56 dependiendo del país.

Dicho todo lo anterior, cabe señalar que otras perspectivas como la «independencia judicial» no se computan porque no pueden ser medidas o bien no generan un conjunto de información que pueda ser utilizado para proponer reformas a la justicia. A pesar de ello, la apuesta por la medida de «lentitud» como aproximación de la calidad o de la eficacia del sistema judicial es también útil en el campo de las percepciones. La lentitud parece guardar también una relación con la confianza en la justicia. La OCDE encontró que un incremento del 10 por 100 en el plazo de resolución de un caso estaba asociado con una reducción de 2 puntos porcentuales en la probabilidad de que los ciudadanos tengan confianza en el sistema. La OCDE estudió esta relación haciendo uso del *World Value Survey* e información del *DB*.

III. LA SITUACIÓN DE LA JUSTICIA ESPAÑOLA

Los estudios mencionados (OCDE, CEPEJ y *DB*) hacen referencia a España y son muy valiosos porque ponen en perspectiva (y con medidas comparables) distintos parámetros del funcionamiento de la justicia. Más abajo se hace referencia a alguno de sus resultados.

Junto a estos trabajos, es posible también hacer mención a información propiamente de España que, si bien no es necesariamente comparable a nivel internacional, permite realizar

un análisis de la situación del sistema judicial español a nivel local (desagregado) y a lo largo del tiempo.

1. Fuentes de información españolas y estudios específicos para España

La OCDE ha encontrado evidencia que apunta a que la producción de información estadística sobre el funcionamiento del sistema judicial (por ejemplo, la duración o el número de casos) y su recopilación está relacionada con un mejor rendimiento judicial.

España posee, de hecho, muy buenas estadísticas judiciales, haciendo públicas algunas series de datos incluso desde 1995.

La «estadística judicial» de la que hablamos es proporcionada por el Consejo General del Poder Judicial (CGPJ) y cubre las dimensiones temporales, por jurisdicción y a nivel de órganos judiciales concretos. Para poder ser utilizada, la información debe ser tratada, pero el CGPJ publica también informes de análisis relativamente amplios basándose en esas estadísticas, como *La justicia dato a dato*.

En parte basados en estas estadísticas, se encuentran los proyectos de investigación referenciados en el cuadro n.º 1 para España. Cabe recordar también otras publicaciones no cubiertas en el cuadro, como el informe completo realizado en 2020 para la Región de Murcia (Mora-Sanguinetti y Martínez-Matute, 2020) o el más general para toda España (Mora-Sanguinetti, 2022).

Otros análisis no cubiertos en el cuadro n.º 1 son los publicados

por Virginia Rosales (Rosales y Jiménez-Rubio [2017] y Rosales-López [2008]). Cabe mencionar análisis más antiguos, como los publicados por Iglesias y Arias (2007) o por Padilla *et al.* (2007). Finalmente, a nivel de obras académicas generales sobre el funcionamiento de la justicia, cabe recordar también los libros de Santos Pastor (1993 y 2016) y Francisco Cabrillo (Cabrillo y Fitzpatrick, 2011).

Además del trabajo de investigadores concretos, diversas organizaciones están desarrollando análisis con información estadística de la situación del sistema judicial español. Cabe destacar a la Asociación Española de Derecho y Economía (AEDE), la Fundación para la Investigación sobre el Derecho y la Empresa (FIDE) o la Fundación Wolters Kluwer.

2. Resultados cuantitativos

Es posible alcanzar numerosos resultados estadísticos que ponen de manifiesto la situación de la justicia en España en términos objetivos.

Antes del inicio de la pandemia por la COVID-19, en 2019, la Justicia española, en todo lo relacionado con conflictos entre empresas y ciudadanos, es decir, de carácter civil, tenía capacidad para resolver solamente 1 de cada 1,8 casos pendientes de su pila de trabajo. Durante la pandemia, en 2020, esa relación empeoró hasta 1 de cada 2.

En general, una medida usual de ineficacia de la justicia es la denominada «tasa de congestión judicial». Siguiendo los distintos resultados expuestos por la bibliografía, una mayor congestión se relaciona con mayores costes

y tiempo para que las empresas vean resueltos sus conflictos por el sistema judicial. El gráfico 1 muestra su evolución en los últimos años.

La tasa se calcula como el cociente de la suma de nuevos casos por resolver que llegan a un órgano judicial concreto más aquellos casos todavía pendientes por resolver, dividida por el número de casos que el órgano judicial es capaz de resolver en el período considerado.

$$\text{Tasa de congestión judicial}_{p,t} = \frac{\text{Casos pendientes}_{p,t-1} + \text{Casos nuevos}_{p,t}}{\text{Casos resueltos}_{p,t}} [1]$$

Si se aproxima la duración (en días) utilizando la metodología de la CEPEJ, se necesitaban 285 días aproximadamente para resolver un proceso contencioso civil en 2019. En 2020 el cálculo alcanzaba ya 354 días.

Estos cálculos no son directamente comparables con los que se pueda hacer para otros países.

CUADRO N.º 2

LA DURACIÓN (EN DÍAS) DE LOS PROCEDIMIENTOS JUDICIALES EN ESPAÑA Y OTROS PAÍSES

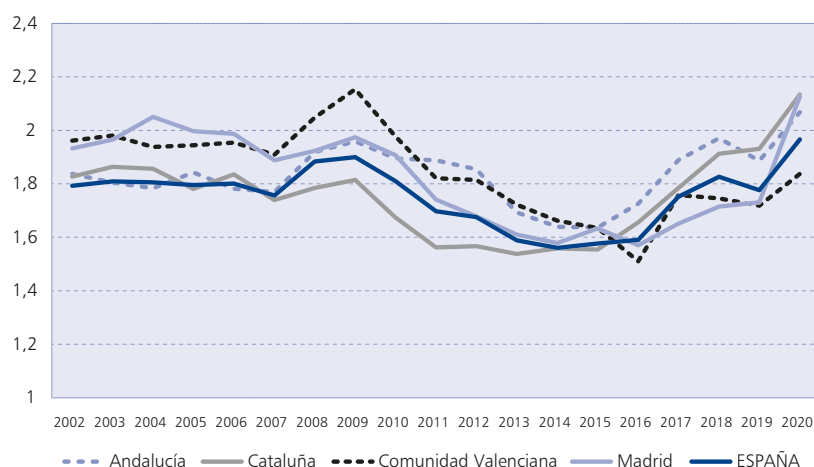
Reino Unido	350			437
España	272	318	282	510
Francia	274	348	353	447
Italia	564	532	514	1120
Alemania	200	198	196	499
Media	332	349	336,3	602,6
Promedio sin Italia	274	288	277	473,3
Fuente	OCDE	CEPEJ	CEPEJ	DB
Año de publicación	2013	2016	2018	2020
Año de análisis	2011	2014	2016	2019

Nota: El Reino Unido está representado por Inglaterra y Gales en la primera columna.
Fuentes: Banco Mundial (2020), CEPEJ (2016 y 2018), Palumbo et al. (2013b).

Una comparación posible de la duración de los procedimientos de forma homogénea a nivel de toda la OCDE es la proporcionada por Palumbo et al. (2013a). Para la jurisdicción civil en primera instancia, el resultado para España fue de 272 días, por debajo de Francia (274) o de Inglaterra y Gales (350), pero por encima

de Alemania (200 días) (véase el cuadro n.º 2). En el cuadro n.º 2 puede encontrarse también información de otras fuentes, como la Comisión Europea para la Eficacia de la Justicia (CEPEJ) y resultados del *Doing Business* (DB). Las distintas metodologías no deben ser comparadas entre sí, si bien las metodologías de la OCDE y de la CEPEJ son similares.

GRÁFICO 1
EVOLUCIÓN DE LA TASA DE CONGESTIÓN JUDICIAL EN ESPAÑA Y EN LAS COMUNIDADES AUTÓNOMAS

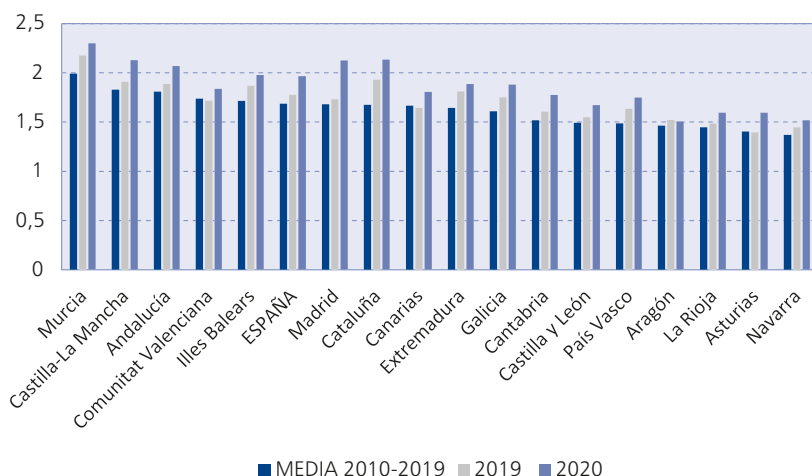


Fuente: Elaboración propia a partir del CGPJ (2021).

A nivel desagregado, España muestra una alta variabilidad local en la tasa de congestión judicial. La congestión es mayor en las provincias de Andalucía, Cataluña y Madrid. Es decir, el problema se agrava en las provincias mediterráneas y en Madrid, que concentran la mayor parte de la población española (véase el gráfico 2).

La ineficacia judicial puede tener implicaciones negativas para la eficiencia económica. Esta sección expone un caso reciente de análisis que explica uno de los motivos por los que la justicia puede tornarse ineficaz. En concreto por el impacto de un marco regulatorio com-

GRÁFICO 2
COMPARATIVA DE LA TASA DE CONGESTIÓN JUDICIAL EN TODAS LAS COMUNIDADES AUTÓNOMAS



Fuente: Elaboración propia a partir del CGPJ (2021).

plejo (véase De Lucio y Mora-Sanguinetti, 2021 y 2022). A continuación, se proporcionan detalles de un caso práctico, de nuevo analizado recientemente, de impacto negativo del sistema judicial en la economía, específicamente en la inversión a nivel de empresa (véanse Dejuán y Mora-Sanguinetti, 2021 y Mora-Sanguinetti, 2021a).

IV. IMPLICACIONES PRÁCTICAS: LA RELACIÓN DE LA JUSTICIA CON LA REGULACIÓN Y EL IMPACTO DE SU FUNCIONAMIENTO EN LA INVERSIÓN EMPRESARIAL

1. El impacto de la complejidad de la regulación en la eficacia judicial

El sistema judicial, como manifestación de las denominadas «instituciones de ejecución»

(siguiendo la clasificación de North), no es una institución aislada. De hecho, tiene como función ejecutar las normas y los contratos. Por tanto, se relaciona de forma estrecha con la regulación de la economía. La normativa es una manifestación de las «instituciones formales».

La «complejidad» de la regulación, que es un concepto relacionado con la forma en la que se regula y no tanto con su contenido, puede tener implicaciones para el funcionamiento de la justicia, lo que a su vez impacta en la economía. La «complejidad» puede venir derivada de que exista un gran volumen de normativa, que la normativa sea lingüísticamente compleja o que guarde muchas relaciones con otras normas (obligando al usuario a consultar un gran número de fuentes legales para entender completamente un conjunto de disposiciones).

La OCDE (Palumbo *et al.*, 2013a) ya había puesto de ma-

nifiesto que una baja «calidad» de la regulación, medida con los indicadores del Banco Mundial (*regulatory quality indicators*), se relacionaba con mayor litigiosidad. Del análisis de la OCDE se extrae, a su vez, que un aumento de la litigación se relaciona con aumentos en los plazos de los procedimientos judiciales, con impactos en los abogados y los servicios legales (Mora-Sanguinetti, 2019).

Para el caso concreto de España, De Lucio y Mora-Sanguinetti (2021 y 2022), analizaron el impacto de la complejidad de las normas de las comunidades autónomas españolas, en concreto toda la normativa con fuerza de ley, en la duración de los procedimientos judiciales en esas regiones (5). Para estudiar la relación entre las dos variables tuvieron en cuenta la mayor o menor presencia de abogados en cada lugar, siguiendo a Carmignani y Giacomelli (2010) y Mora-Sanguinetti y Garoupa (2015) (6) y una medida de productividad. Los resultados indican que una mayor complejidad «relacional», es decir que las normas hagan referencia a muchas otras normas para poder completar su contenido, dificulta o enlentece la acción judicial, en concreto en la jurisdicción civil (aquella relacionada con la aplicación de los contratos y la actividad privada de empresas y ciudadanos).

2. El impacto de la eficacia judicial en la inversión empresarial

Precisamente, a nivel práctico, un incremento de la ineficacia judicial en su jurisdicción civil tiene implicaciones negativas para la inversión a nivel de empresa en España. Cabe desarrollar los

argumentos esgrimidos en la investigación de Dejuán y Mora-Sanguinetti (2021), Mora-Sanguinetti (2021a) que, de hecho, fueron detallados en esta misma revista recientemente (véase Mora-Sanguinetti, 2021b).

El análisis se realizó recopilando información de las decisiones de inversión a nivel de empresa en España (en concreto para 653.289 empresas) durante el período 2002-2016, a partir de la Central de Balances Integrada (CBI) del Banco de España.

Esta información se relacionó con una tasa de congestión judicial para la jurisdicción civil calculada a nivel provincial. Como se indicaba antes, existen en España unas diferencias importantes en la eficacia judicial a nivel desagregado, lo que permite explotar esa información para conocer sus impactos en la inversión.

Por supuesto, otra información fue tenida en cuenta, como los pagos de intereses de las empresas, su volumen de deuda y otros costes financieros.

En términos cuantitativos, un aumento en una unidad en la tasa de congestión judicial incita una caída de la inversión de 1 punto porcentual. Cabe ejemplificar los resultados hablando de provincias concretas. En 2010, Alicante era una de las provincias con mayor nivel de congestión, 2,2 (220 casos no resueltos por cada 100 resueltos), mientras que Álava presentaba uno de los niveles más bajos, 1,4 (140 casos no resueltos por cada 100 resueltos). De forma coherente con los resultados, si Alicante hubiera tenido la eficacia judicial de Álava, su ratio de inversión empresarial habría sido 0,8 puntos porcentuales superior.

Bajando a mayor detalle, los resultados apuntan a que el impacto de la ineficacia judicial es tres veces mayor en las empresas de mayor tamaño que en las pequeñas, por lo que cabe recordar que el tamaño de empresa es relevante para entender la evolución del crecimiento de la productividad. Siguiendo a López-García y Montero (2012), las empresas pequeñas están relacionadas con menores probabilidades de innovar.

V. CONCLUSIONES

La evaluación y el análisis de la justicia y su funcionamiento parecen esenciales para una sociedad desarrollada, ya sea desde un punto de vista jurídico o desde un punto de vista económico. Por el lado jurídico, el sistema judicial es un instrumento fundamental de garantía de derechos y de libertades.

Esta garantía se manifiesta en dos vertientes distintas: en primer lugar, porque los ciudadanos y las empresas se enfrentan, en el ámbito privado, a otros sujetos (privados) que resultan ser oportunistas e incumplen sus obligaciones. El sistema judicial actuaría en ese ámbito como disciplinante de esas relaciones. En segundo lugar, esos mismos ciudadanos o empresas pueden verse enfrentados a la actuación incorrecta por parte de los poderes públicos. El sistema judicial es, en ese caso, una barrera o límite contra la violación de los derechos de propiedad por parte de la acción gubernamental.

Precisamente, esas dos perspectivas (la privada y la pública) tienen una alta relevancia desde el punto de vista del análisis económico y ya fueron puestas en valor por Coase (1960) y Acemoglu y Johnson (2005).

La calidad de la justicia se relaciona, en fin, con una mayor seguridad jurídica, lo que sostiene la inversión (Dejuán y Mora-Sanguinetti, 2021), una más fácil combinación de factores productivos y la toma de decisiones empresariales. Vista de forma global, la acción de la justicia determina parcialmente el comportamiento de la productividad. Además, de forma específica, más allá de su impacto en las decisiones de inversión, la ineficacia de la justicia afecta, por ejemplo, a los mercados de crédito o al mercado de vivienda (reduciendo el peso del mercado de alquiler).

Una vez establecida la importancia de la cuestión, este artículo discute cómo proporcionar una medición de la eficacia (o ineficacia) del sistema judicial. En general, la bibliografía a nivel internacional reduce la evaluación a tres parámetros: la medición de la lentitud de su funcionamiento, su coste o su predictibilidad. Esas tres medidas, sin embargo, ceden ante la primera: la lentitud de la justicia parece ser el indicador global más adecuado de calidad o eficacia.

A continuación, esta investigación repasa las fuentes de información estadística que pueden ser útiles para realizar la evaluación: a nivel internacional (cubriendo también el caso español) destacan los informes y bases de datos de la OCDE, la CEPEJ y el proyecto *Doing Business* del Banco Mundial (esta última discontinuada). A nivel nacional, España por su lado posee, de hecho, muy buenas estadísticas judiciales, haciendo públicas algunas series de datos desde 1995.

Con esa información es posible alcanzar numerosos resul-

tados estadísticos que ponen de manifiesto la situación de la justicia en España en términos objetivos. Antes del inicio de la pandemia por la COVID-19, en 2019, la justicia española, en todo lo relacionado con conflictos entre empresas y ciudadanos, es decir, de carácter civil, tenía capacidad para resolver solamente 1 de cada 1,8 casos pendientes de su pila de trabajo. Durante la pandemia, en 2020, esa relación empeoró hasta una proporción de 1 de cada 2. Junto a ese resultado general, España muestra una alta variabilidad local en la tasa de congestión judicial. La congestión es mayor en las provincias de Andalucía, Cataluña y Madrid. Por tanto, cabe señalar que el problema se agrava en las provincias mediterráneas y en Madrid, que concentran la mayor parte de la población española.

Finalmente, el artículo desarrolla algunas implicaciones prácticas y ejemplos (en cuanto a los impactos económicos) del funcionamiento de la justicia en España. En primer lugar, se analiza uno de los motivos por los que la justicia puede tornarse ineficaz y este es el de encontrarse un marco regulatorio complejo. Los resultados apuntan a que una complejidad «relacional» más alta, en otras palabras que las normas realicen referencias a muchas otras normas para lograr completar su contenido, dificulta o enlentece la acción judicial. En segundo lugar, se proporcionan detalles de la relación entre justicia e inversión, poniendo de manifiesto que un incremento de la ineficacia judicial en la jurisdicción civil tiene implicaciones negativas para la inversión a nivel de empresa en España.

NOTAS

(*) Las opiniones y las conclusiones recogidas en este artículo representan las ideas del autor, con las que no necesariamente tiene que coincidir el Banco de España o el Eurosistema.

(1) Regulados en los artículos 24.1 de la Constitución Española y en el 6.1 del Convenio Europeo de Derechos Humanos.

(2) Como se discute en MORA-SANGUINETTI (2010a), el indicador del DB («enforcing contracts») se inspira en el trabajo de DJANKOV *et al.* (2003). DJANKOV evalúa negativamente el denominado «formalismo» y toma como modelo el *neighbourhood model* del «Derecho Común». De hecho, los países de ese sistema legal obtienen mejores resultados tanto en el trabajo de DJANKOV como en el DB.

(3) En esta sección la metodología que se utiliza de ejemplo es la de la OCDE (PALUMBO *et al.*, 2013a y 2013b).

(4) El debate que generan estas cuestiones es más amplio. Como ya se vio más arriba, litigar es un derecho. Sin embargo, el exceso de litigación implica en la práctica la anulación de ese derecho por el propio colapso de la justicia (aumentando de forma relevante la lentitud). Distintos mecanismos han sido discutidos para lograr conjugar ambas perspectivas (permitir la litigación, racionalizando el uso del sistema judicial). Uno de los instrumentos ha sido la aprobación de sistema de «tasas judiciales» (véase MORA-SANGUINETTI y MARTÍNEZ-MATUTE [2019 y 2022]). El análisis de estas cuestiones supera el objetivo de este artículo.

(5) La duración es una aproximación a la congestión del sistema judicial.

(6) Cabe recordar que MORA-SANGUINETTI y GAROUPA (2015) incluyeron en su análisis del mercado de la abogacía el número de asociaciones y el número de empresas, en términos per cápita, para tener en cuenta las diferencias entre distintas fuentes de litigación. Cabe esperar que contextos o ambientes con una mayor «complejidad» social sean más proclives a la litigación.

BIBLIOGRAFÍA

- ACEMOGLU, D. y JOHNSON, S. (2005). Unbundling Institutions. *Journal of Political Economy*, 113(5), pp. 949-995.
- ARDAGNA, S. y LUSARDI, A. (2010). The international differences in entrepreneurship. En J. LERNER y A. SCHOAR (eds.), *Explaining international differences in entrepreneurship: The role of individual characteristics and regulatory constraints*. Chicago: University of Chicago Press.

BANCO MUNDIAL (2020). *Doing Business 2020. Comparing Business Regulation in 190 Economies*. Washington, DC.: International Bank for Reconstruction and Development/The World Bank.

BECK, T., DEMIRGUC-KUNT, A. y MAKSIMOVIC, V. (2006). The Influence of Financial and Legal Institutions on Firm Size. *Journal of Banking & Finance*, 30(11), pp. 2995-3015.

BERGER, H. y NEUGART, M. (2011). Labour Courts, Nomination Bias, and Unemployment in Germany. *European Journal of Political Economy*, 27, pp. 659-673.

CARMIGNANI, A. y GIACOMELLI, S. (2010). Too many lawyers? Litigation in Italian civil courts. Temi di Discussione. *Working Papers*, n.º 745. Banca d'Italia.

CASAS-ARCE, P. y SAIZ, A. (2010). Owning versus Renting: Do Courts Matter? *Journal of Law & Economics*, 53(1), pp. 137-165.

CEPEJ - EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (2016). *European judicial systems. Efficiency and quality of justice. Edition 2016 (2014 data)*. CEPEJ Studies n.º 23.

CEPEJ - EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (2018). *European judicial systems. Efficiency and quality of justice. Edition 2018 (2016 data)*. CEPEJ Studies, n.º 26.

CHEMIN, M. (2012). Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India. *Journal of Law, Economics, & Organisation*, 28(3). pp. 460-485.

COASE, R. H. (1960). The problem of social cost. *The Journal of Law & Economics*, 3, pp. 1.44.

DE LUCIO, J. y MORA-SANGUINETTI, J. S. (2021). New dimensions of regulatory complexity and their economic cost. An analysis using text mining. *Working Paper*, 2107. Banco de España.

DE LUCIO, J. y MORA-SANGUINETTI, J. S. (2022). Drafting «better regulation»: the economic cost of regulatory complexity. *Journal of Policy Modeling* (de próxima publicación). doi: <https://doi.org/10.1016/j.jpolmod.2021.10.003>

DEJUÁN, D. y MORA-SANGUINETTI, J. S. (2021). Which legal procedure affects business investment most, and which companies are most sensitive? Evidence from microdata. <i>Economic Modelling</i>, 94, pp. 201-220. DESAI, M., GOMPERS, P. y LERNER, J. (2005). Institutions, capital constraints and entrepreneurial firm dynamics: Evidence from Europe. <i>Harvard NOM Working Paper</i>, n.º 03-59. DJANKOV, S., LA PORTA, R., LÓPEZ DE SILANES, F. y SHLEIFER, A. (2003). Courts. <i>The Quarterly Journal of Economics</i>, 118, pp. 453-517. FABBRI, D. (2010). Law enforcement and firm financing: Theory and evidence. <i>Journal of the European Economic Association</i>, 8(4), pp. 776-816. FABBRI, D. y PADULA, M. (2004). Does poor legal enforcement make households credit-constrained? <i>Journal of Banking & Finance</i>, 28(10), pp. 2369-2397. GARCÍA-POSADA, M. y MORA-SANGUINETTI, J. S. (2015a). Does (average) size matter? Court enforcement, business demography and firm growth. <i>Small Business Economics</i>, 44(3), pp. 639-669. GARCÍA-POSADA, M. y MORA-SANGUINETTI, J. S. (2015b). Entrepreneurship and Enforcement Institutions: Disaggregated Evidence for Spain. <i>European Journal of Law & Economics</i>, 40(1), pp. 49-74. GIACOMELLI, S. y MENON, C. (2017). Does weak contract enforcement affect firm size? Evidence from the neighbour's court. <i>Journal of Economic Geography</i>, 17(6), pp. 1251-1282. IGLESIAS, P. y ARIAS, X. C. (2007). El sistema judicial español: una revisión de los principales problemas de la oferta y demanda de tutela judicial. <i>Presupuesto y Gasto Público</i>, 47, pp. 137-160. JAPPELLI, T., PAGANO, M. y BIANCO, M. (2005). Courts and Banks: Effects of Judicial Enforcement on Credit Markets. <i>Journal of Money Credit and Banking</i>, 37(2), pp. 223-244.	JIMENO, J. F., MARTÍNEZ-MATUTE, M. y MORA-SANGUINETTI, J. S. (2020). Employment Protection Legislation, Labor Courts, and Effective Firing Costs. <i>IZA Journal of Labor Economics</i>, 9(1). doi: https://doi.org/10.2478/izajole-2020-0002 LAEVEN, L. y C. WOODRUFF (2007). The Quality of the Legal System, Firm Ownership and Firm Size. <i>The Review of Economics and Statistics</i>, 89, pp. 601-614. MARINESCU, I. (2011). Are Judges Sensitive to Economic Conditions? Evidence from UK Employment Tribunals. <i>Industrial and Labour Relations Review</i>, 64(4), pp. 673-698. MORA-SANGUINETTI, J. S. (2010a). A Characterization of the Judicial System in Spain: Analysis with Formalism Indices. <i>Economic Analysis of Law Review</i>, 1(2), pp. 210-240. MORA-SANGUINETTI, J. S. (2010b). The effect of institutions on the European housing markets: an economic analysis. <i>Estudios Económicos</i>, 77. Madrid: Banco de España. MORA-SANGUINETTI, J. S. (2012). Is judicial inefficacy increasing the weight of the house property market in Spain? Evidence at the local level. <i>SERIEs, Journal of the Spanish Economic Association</i>, 3(3), pp. 339-365. MORA-SANGUINETTI, J. S. (2019). Producción normativa, mercado de la abogacía y litigación en España. <i>Dossier, Diciembre/2019</i>. Thomson Reuters. MORA-SANGUINETTI, J. S. (2020). Los impactos económicos de la justicia en España. <i>Informe del Observatorio de la Actividad de la Justicia</i>. Fundación Wolters Kluwer. https://tienda.wolterskluwer.es/pl/observatorio-de-la-actividad-de-la-justicia-informe-2020 MORA-SANGUINETTI, J. S. (2021a). El impacto de la eficacia de la justicia en la inversión empresarial en España. Artículos Analíticos. <i>Boletín Económico-Banco de España</i>, 3/2021. MORA-SANGUINETTI, J. S. (2021b). Justicia y economía: la eficiencia del sistema judicial en España y sus impactos económicos. <i>Papeles de Economía Española</i>, 168.	MORA-SANGUINETTI, J. S. (2022). <i>La factura de la injusticia. Sistema judicial, economía y prosperidad en España</i>. Madrid: Tecnos. MORA-SANGUINETTI, J. S. y GAROUPA, N. (2015). Do lawyers induce litigation? Evidence from Spain, 2001-2010. <i>International Review of Law & Economics</i>, 44, pp. 29-41. MORA-SANGUINETTI, J. S. y MARTÍNEZ-MATUTE, M. (2019). An economic analysis of court fees: evidence from the Spanish civil jurisdiction. <i>European Journal of Law & Economics</i>, 47(3), pp. 321-359. MORA-SANGUINETTI, J. S. y MARTÍNEZ-MATUTE, M. (2020). <i>Los impactos económicos del funcionamiento de la justicia en la Región de Murcia</i>. Consejo Económico y Social de la Región de Murcia. Murcia. MORA-SANGUINETTI, J. S., MARTÍNEZ-MATUTE, M. y GARCÍA-POSADA, M. (2017). Credit, crisis and contract enforcement: evidence from the Spanish loan market. <i>European Journal of Law & Economics</i> 44(2), pp. 361-383. NORTH, D. C. (1990). <i>Institutions, institutional change and economic performance</i>. Cambridge: Cambridge University Press. NUNN, N. (2007). Relationship-Specificity, Incomplete Contracts and the Pattern of Trade. <i>The Quarterly Journal of Economics</i>, 122(2), pp. 569-600. PADILLA, J., LLORENS, V., PEREIRAS, S. y WATSON, N. (2007). Eficiencia judicial y eficiencia económica: el mercado crediticio español. En <i>La Administración Pública que España necesita. Libro Marrón</i>. Madrid: Círculo de Empresarios. PALUMBO, G., GIUPPONI, G., NUNZIATA, L. y MORA-SANGUINETTI, J. S. (2013a). The Economics of Civil Justice: New Cross-Country Data and Empirics. <i>Working Papers</i>, n.º 1060. OECD Economics Department. PALUMBO, G., GIUPPONI, G., NUNZIATA, L. y MORA-SANGUINETTI, J. S. (2013b). Judicial performance and its determinants: a cross-country perspective. <i>OECD Economic Policy Paper</i>, n.º 5.
--	--	---

ROSALES, V. y JIMÉNEZ-RUBIO, D. (2017). Empirical analysis of civil litigation determinants: The Case of Spain. <i>European Journal of Law & Economics</i>, 44(2), pp. 321-338. ROSALES-LÓPEZ, V. (2008). Economics of Court Performance: An Empirical	Analysis. <i>European Journal of Law & Economics</i>, 25(3), pp. 231-251. SHVETS, J. (2013). Judicial Institutions and Firms' External Finance: Evidence from Russia. <i>The Journal of Law,</i>	<i>Economics, and Organization</i> 29(4), pp. 735-764. STEPHEN, F., URBANO, D. y VAN HEMMEN, S. (2009). The responsiveness of entrepreneurs to working time regulations. <i>Small Business Economics</i>, 32(3), pp. 259-276.
--	--	---

Resumen

Se revisan datos y métodos para evaluar las políticas sanitarias y de salud, de enorme trascendencia económica durante la pandemia, y la institucionalización de la evaluación en España. Durante la pandemia, el gasto sanitario es solo una pequeña parte del coste total de la COVID. La prevención es altamente coste-efectiva. Proponemos áreas para evaluación: digitalización en sanidad, gestión indirecta de la gestión sanitaria, tecnologías de alto coste, intervenciones económicas y fiscales sobre la salud de la población, modelo MUFACE, y las consecuencias a largo plazo de la COVID-19 sobre la salud y sobre la desigualdad.

Palabras clave: evaluación de políticas, sanidad, análisis coste-efectividad, evaluación económica.

Abstract

We review data and methods for evaluating health and health policies, of enormous economic significance during the pandemic, and the institutionalization of evaluation in Spain. During the pandemic, health expenditure is only a small part of the total cost of COVID. Prevention is highly cost-effective. We propose areas for evaluation: digitalization in healthcare, indirect management of healthcare management, high-cost technologies, economic and fiscal interventions on population health, MUFACE model, and the long-term consequences of COVID-19 on health and inequality.

Keywords: policy evaluation, healthcare, cost-effectiveness analysis, economic evaluation.

JEL classification: H51, I14, I18.

EVALUACIÓN DE POLÍTICAS. SALUD Y SANIDAD

Beatriz GONZÁLEZ LÓPEZ-VALCÁRCEL (*)

Universidad de Las Palmas de Gran Canaria

I. INTRODUCCIÓN

EN los últimos dos años ha habido enormes avances en el conocimiento humano, con la ciencia como protagonista. Una enfermedad infecciosa ha marcado el paso a la economía mundial y específicamente al sector sanitario, que se ha vuelto más visible y pieza esencial en la gestión de las nuevas crisis, a la vez sanitarias y económicas.

El mayor reto de los países durante estos dos años ha consistido en encontrar soluciones para compatibilizar la pandemia con la vida de las personas y mantener el tejido productivo y el crecimiento económico a medio plazo. Contar con una sanidad bien organizada, reactiva, flexible y con los recursos necesarios es un requisito. La sanidad ya tenía su lista de problemas y retos pendientes, pero la agenda varió sustancialmente con la pandemia. Nunca la sanidad y la economía habían sido tan interdependientes, nunca las políticas sanitarias habían tenido tanto de políticas económicas. La pregunta del millón es si las decisiones que se tomaron en su momento consiguieron minimizar el impacto de la pandemia en los distintos frentes.

Articular una respuesta rigurosa a esa pregunta no puede hacerse sin desarrollar previamente un proceso de evaluación completo y metodológicamente robusto. La evaluación de polí-

ticas en general, y desde luego también la gestión de la pandemia, requiere datos accesibles y de calidad y el uso de métodos reproducibles (Casado y Pino, 2021). La disponibilidad global de datos de grano fino, sobre todo epidemiológicos, pero también económicos, de movilidad humana o de la gestión pública de la pandemia es cada vez mayor. Paralelamente, ha habido un enorme desarrollo y presencia de la *ciencia de datos*, agnóstica y omnipresente, al albur de la mejora exponencial en disponibilidad y acceso a algunos tipos de datos. La inteligencia artificial (IA) se ha convertido en objetivo de los Gobiernos, tal vez más como un fin en sí mismo que como un medio para avanzar en el conocimiento y el bienestar de la humanidad.

En todo caso, y más allá del caso particular de la respuesta a la pandemia de COVID-19, la evaluación de políticas de salud es particularmente compleja. Aunque es habitual que se pierda de vista, la salud, objetivo último, es más que sanidad, y se genera fuera del sector sanitario mediante políticas intersectoriales, de vivienda y urbanismo, laborales y fiscales, entre otras. Además, la evaluación de las políticas de salud está menos incentivada que, por ejemplo, la evaluación de tecnologías sanitarias como son los medicamentos o los procedimientos quirúrgicos, por lo que suele concentrar menos recursos y esfuerzos.

En este artículo se reflexiona sobre los datos y métodos para evaluar las políticas sanitarias y de salud y sobre la organización de dicha evaluación en España. Tras esta introducción, la sección segunda de este artículo se centra en los *datos* necesarios para evaluar los avances que se han obtenido en los últimos años y los problemas que han surgido. En la tercera sección se abordan los cambios en los *métodos* de evaluación, en parte relacionados con las tendencias en la disponibilidad de datos. La sección cuarta aborda específicamente la modelización epidemiológica y económica de la pandemia, que aporta *inputs* imprescindibles para la evaluación de las políticas públicas relacionadas con la COVID-19. Relacionado con esa última sección, la quinta sección considera la evaluación de la *macrogestión* de la pandemia, especialmente en España. La sexta sección se pregunta si los *métodos clásicos de evaluación económica* en sanidad siguen siendo aplicables o deben ser reformulados y adaptados, y se analizan los cambios recientes en la arquitectura institucional de la evaluación de tecnologías sanitarias en nuestro país. También se presentan algunas aplicaciones específicas de evaluación económica relacionada con la COVID-19. La sección séptima abre el foco, presentando ejemplos relevantes de evaluación de políticas públicas de salud y sanidad que no están relacionadas con la pandemia, y se apuntan temas relevantes para una agenda priorizada de evaluación de políticas en el ámbito sanitario y de la salud. Finalmente, se concluye con el resumen de los principales mensajes y prescripciones.

1. Los datos necesarios para evaluar

La materia prima esencial para evaluar las políticas de salud y sanitarias son los datos. La evaluación de políticas en la actualidad se ve influida por tres grandes tendencias:

- Disponibilidad de *registros sanitarios* poblacionales en lugar de muestras. Esos datos se crean en formato digital y por tanto están o pueden estar accesibles *online*. Hay dos retos, hacerlos interoperables y combinar bases de datos individuales de distintos registros, preservando la confidencialidad y el anonimato. Las historias clínicas combinadas con bases de datos de dispensaciones farmacéuticas y de imagen tienen un enorme potencial para la investigación observacional con datos en el mundo real (*real world data*). Algunas comunidades autónomas (CC. AA.) van por delante y ya cuentan con registros poblacionales de datos de salud, con usos en investigación evaluativa (García-Sempere *et al.*, 2020).
- Disponibilidad de *datos longitudinales*, que permiten controlar por heterogeneidad no observable entre personas –pacientes, proveedores– y organizaciones sanitarias. Es el caso de las historias clínicas, pero también de los datos de utilización, actividad y calidad de los centros sanitarios a nivel de mesogestión, como los que publica en abierto el Observatorio de Resultados del Servicio Madrileño de Salud (<https://www.comunidad.madrid/servicios/salud/observatorio-resultados-servicio-madrileno-salud> o AQUAS

en Cataluña (<https://aguas.gencat.cat/ca/inici>)

- Disponibilidad creciente de *big data* (vs. datos estructurados), caracterizados por las cinco «uves»: volumen, variedad, velocidad, veracidad y valor. Representan un reto para los modelos econométricos y estadísticos tradicionales y son *inputs* esenciales para la *ciencia de datos*.

El contar con información digitalmente accesible facilita la evaluación y la investigación evaluativa. Todavía no podemos valorar hasta qué punto las políticas de transparencia activa mediante la incorporación de estándares de datos abiertos y de la Carta Internacional de Datos Abiertos (IODC) (Ubaldi, 2013) han avanzado durante los años de pandemia. El índice OURdata mide la disponibilidad, accesibilidad y reutilización de los datos del Gobierno. Según la OCDE, se trata de «un bien público que debería ser provisto con un propósito definido, proactivamente y con foco en la reutilización, en línea con las necesidades del usuario» (OCDE, 2020). El *Digital Government Index* de la OCDE incorpora las seis dimensiones del *Digital Government Policy Framework*: 1) digital desde el diseño; 2) sector público impulsado por los datos; 3) Gobierno como plataforma; 4) abierto por defecto; 5) enfocado al usuario; y 6) proactividad. En 2019, último año para el que hay datos, España ocupaba el séptimo puesto de los 33 países de la OCDE en el índice global. Aunque destaca entre los primeros puestos (puesto 4 de 33) en las dimensiones 1 (administración de diseño digital), 2 (un sector público impulsado por los datos) y 6 (proactivo), hay mucho mar-

gen de mejora en el enfoque hacia el usuario (puesto 12), y sobre todo en el establecimiento del principio de accesibilidad del público a los datos por defecto, en el que ocupamos el puesto 23, a la cola.

En España se han dado algunos pasos adelante, como el acuerdo de abril de 2021 entre el INE, la Agencia Estatal de Administración Tributaria, la Seguridad Social, el Servicio Público de Empleo Estatal y el Banco de España para facilitar datos anonimizados cruzados para investigación. También las líneas estratégicas del Programa anual 2021 del Plan Estadístico Nacional 2021-2024 (Real Decreto 150/2021) (Póo y Galiano, 2021) y la estadística experimental del INE (<https://www.ine.es/experimental/experimental.htm>) representan avances en la disponibilidad de datos *de grano fino* casi en tiempo real. La estadística experimental incluye desde la información semanal de movilidad humana, cotidiana y estacional, obtenida de las antenas de teléfonos móviles, que permite predecir mejor la dinámica de la pandemia, hasta la estimación del número de defunciones semanales durante el brote de COVID-19: (https://www.ine.es/experimental/defunciones/experimental_edes_proyecto.pdf).

2. Cambios en los métodos de evaluación

«Para quien solo tiene un martillo, todo son clavos». Frente al enorme avance en la disponibilidad de datos, la metodología para procesarlos con fines de evaluación de políticas públicas está relativamente estancada. Desde la economía, los enfoques estructural (Heckman, 2001) y la revolución de la credibilidad

(Angrist y Pischke, 2010), que le valieron el Nobel de Economía en 2000 a Heckman y en 2021 a Angrist, Imbens y Card, tienen ya décadas. La disponibilidad omnipresente de datos masivos resulta en una tendencia a explorar microcausalidades locales más que relaciones generales de causa-efecto, con riesgo de sobreutilización de IA, que está resultando más útil para unas funciones (por ejemplo, para diagnosticar COVID-19 adiestrando algoritmos para que exploren placas de tórax) que para otras (por ejemplo, para contestar preguntas específicas sobre el impacto de determinadas políticas en la utilización sanitaria o en la salud de la población).

El ensayo clínico aleatorizado y doble ciego es el *gold standard* en diseño de datos experimentales en biomedicina, y durante años desde la economía hemos buscado buenos *proxies* cuando no es posible hacer experimentos sociales, cuasiexperimentos o contar con experimentos naturales. Con todo, el número de (auténticos) experimentos sociales relacionados con la salud y con las políticas de salud es abundante, sobre todo en países en desarrollo. El Banco Mundial ha seguido una estrategia de evaluación de impacto de las políticas (Gertler *et al.*, 2016) en la que el diseño e implementación de experimentos juega un papel central. Incluso ha avanzado en el diseño de nuevos métodos para minimizar los sesgos provocados por su propio *staff*. Es lo que se ha dado en llamar *práctica reflexiva* (McKee y Stuckler, 2015), recomendable para otras organizaciones que han de evaluar políticas de salud.

Pero en economía, en ciencias sociales y en los análisis del

comportamiento humano relacionado con la salud, los pequeños ensayos, aunque estén bien diseñados, tienen limitaciones. Por ese motivo, se está propugnando como un avance sustancial la implementación de *megaensayos* para establecer relaciones causa-efecto de distintas intervenciones sobre el comportamiento humano (Royer, 2021). Por ejemplo, un experimento masivo (Milkman *et al.*, 2021) estimó y comparó el efecto de 54 intervenciones diferentes y simultáneas, de cuatro semanas de duración, en la misma población (los 61.293 usuarios de una red de gimnasios en EE. UU.) utilizando la misma definición y medida de resultado (frecuencia de asistencia al gimnasio). Un total de 30 científicos de 15 universidades trabajaron en pequeños equipos independientes en el diseño de las intervenciones. Este megastudio supera las limitaciones de los pequeños ensayos, cuyos resultados solo tienen validez local y son difícilmente comparables entre experimentos.

Pero el tamaño no lo es todo, siendo más importante evitar sesgos por falta de representatividad de las muestras que conseguir tamaños muestrales masivos. Un ejemplo contrastado de la *paradoja del big data* es la estimación del porcentaje de personas vacunadas de COVID en EE. UU. entre enero y mayo de 2021 a partir de grandes encuestas en Facebook (n=250.000 por semana) y del *Census Household Pulse* (n=75.000 cada dos semanas). Con los datos reales de vacunaciones se puso en evidencia que ambas encuestas sobreestimaban el porcentaje de vacunados, en 17 puntos la de Facebook y en 14 puntos la otra (Bradley *et al.*, 2021). Un buen diseño muestral con una muestra 500 veces menor habría

proporcionado estimadores con error cuadrático medio menor y sin sesgo.

II. LOS MODELOS: DE LA EPIDEMIOLOGÍA A LA EPIECONOMÍA

Desde marzo de 2020, ha habido una *pandemia de modelos* COVID (1). La mayor parte han ido desapareciendo, en un proceso darwiniano en el que sobreviven los más capaces de aportar. Algunos han sido adoptados por las instituciones internacionales, como es el caso del modelo de la Universidad Politécnica de Cataluña por el European Centre for Disease Prevention and Control (ECDC) (Català *et al.*, 2021).

Los modelos epidemiológicos más simples partían de los modelos SIR (susceptibles, infectados y recuperados); SEIR (susceptibles, expuestos, infectados y recuperados) del siglo pasado. Son modelos *compartimentales* en los que cada persona pertenece a un grupo de un conjunto exhaustivo y mutuamente excluyente: susceptibles, expuestos, infectados, recuperados. Han ido añadiendo *complejidad* progresivamente, tendiendo hacia modelos de grano fino que incorporan cada vez más información micro sobre movilidad e interacciones sociales, y sobre políticas públicas (restricciones a la movilidad y a la actividad económica) e intentan contestar preguntas sobre efectos de las políticas e intervenciones públicas. Las aplicaciones para la actual pandemia estiman los efectos de la movilidad y las restricciones de los Gobiernos, sobre todo en relación a viajes y movimientos transfronterizos (Chinazzi *et al.*, 2020; Linka *et al.*, 2020; Poletto, Scarpino y Volz, 2020). No es que estos modelos epidémicos espaciales sean nue-

vos (Balcan *et al.*, 2010), pero nunca antes había habido tantos datos para utilizarlos.

Una función importante de los modelos es predecir la presión sobre el uso de los recursos y servicios sanitarios y el déficit de recursos disponibles para otros problemas de salud. Complementariamente al tradicional *forecasting* o predicción de futuro, ha surgido el *nowcasting*, (literalmente «predicción del ahora»), predicción a ultracorto plazo, o más bien análisis de lo que está pasando en tiempo real. Se había aplicado en meteorología, física y otras ciencias, y ahora se han encontrado aplicaciones para tomar decisiones exprés relativas a la pandemia (Rossman y Segal, 2022) (Wu *et al.*, 2021).

Estos modelos utilizan datos masivos *compartidos en tiempo real*, de acceso abierto. Hay proveedores universales, como la John Hopkins y Dataworld. La comisión COVID-19 de *The Lancet* (<https://data.covid19commission.org/>) tiene una base de datos de acceso abierto que actualiza semanalmente, con 154 indicadores para 244 países, más de 10 millones de datos en total.

Otros *inputs* para los modelos eieconómicos son los datos de movilidad (a nivel regional, las visitas diarias a tiendas y ocio, supermercados, parques, farmacias, estaciones de transporte, lugares de trabajo y zonas residenciales) que aporta Google como *inputs* para los modelos, obtenidos del seguimiento que hace Google Maps de los desplazamientos (<https://www.google.com/covid19/mobility/>).

Tanto los datos como los modelos se han globalizado. Hace apenas unos años hubiera sido

impensable que una organización como el Institute of Health Metrics and Evaluation (IHME) de la universidad de Seattle (EE. UU.) ofreciera predicciones regionales sobre la evolución de la pandemia bajo escenarios alternativos de uso de mascarillas y vacunación, no solo para España (<https://covid19.healthdata.org/spain?view=cumulative-deaths&tab=trend>), sino también para cada una de sus CC. AA.

Las reacciones de los Gobiernos frente a la evolución de la pandemia, las medidas que han ido tomando para *aplanar la curva* o para crear *zonas verdes* libres de virus intercomunicadas entre sí, se plasman y cuantifican en índices estandarizados, comparables internacionalmente (Wong *et al.*, 2020). Uno de los más conocidos y utilizados, también por la OCDE, es el *Stringency Index* de Oxford (Hale *et al.*, 2021), que contiene 23 indicadores de respuesta de los Gobiernos: restricciones, instrumentos económicos y respuestas desde el sistema sanitario.

Los modelos son para predecir a corto plazo, pero también para evaluar las políticas de gestión de la pandemia. Utilizando el método científico, se trata de aprender de los datos para explicar qué ha pasado y por qué (y no repetirlo), detectar tempranamente brotes (anticipación) y aconsejar sobre decisiones políticas. Asimismo, los modelos permiten contestar preguntas específicas y encontrar asociaciones. Sin embargo, tienen problemas por la incertidumbre intrínseca de una enfermedad que está en continua evolución. Los modelos de simulación de políticas tienen mayor incertidumbre que los predictivos, en parte debido a que necesitan *inputs* sobre la adherencia de la población a las

restricciones (el propio rebaño es quien determina su inmunidad). No es de extrañar que incluso los modelos más simples, epidemiológicos predictivos, obtengan buenos resultados en el corto plazo (y en el hipercorto, con el *nowcasting*), pero sus errores de predicción a medio plazo (varias semanas o meses) sean muy altos (para un *ranking* de capacidad predictiva de los modelos, véase Friedman *et al.*, 2021). Para Europa, también hay disponible un sitio web que compara predicciones de distintos modelos (2) supervivientes de aquel proceso darwiniano al que se hacía referencia al principio de esta sección.

La modelización ha logrado grandes progresos, pero todavía falta integrar mejor los modelos epidemiológicos y económicos, así como incorporar la dimensión de las *desigualdades* interpersonales y los efectos sobre la desigualdad.

1. La necesidad de evaluar la macrogestión de la pandemia

A grandes rasgos, la gestión de la pandemia abarca: 1) las medidas para limitar los contagios; 2) la disposición y organización de recursos y servicios socio-sanitarios para prevenir y tratar la enfermedad; y 3) las medidas de carácter económico para minimizar los efectos de la pandemia sobre el tejido productivo y sobre las economías familiares.

Las decisiones de los tres tipos se han ido tomando con urgencia y en el ambiente de incertidumbre de cada momento. Conceptualmente, una evaluación *ex ante* ayudaría a decidir el curso de la acción, comparando los efectos de escenarios alternativos correspondientes a distintas

medidas. Las evaluaciones *ex post* permiten valorar *a posteriori* el acierto o la falta de acierto, las buenas y malas decisiones del pasado, para no reincidir cuando llegue la próxima oleada o pandemia.

En puridad, una evaluación de la (macro) gestión de la pandemia debería considerar las dos grandes alternativas de los Gobiernos, la *estrategia de supresión*, con zonas verdes en las que no hay casos de infección, que están intercomunicadas y aisladas de las zonas *rojas*, y la *estrategia de mitigación*, basada en restricciones tanto más serias cuanto peor es la incidencia y otros indicadores de los semáforos COVID (ocupación de camas, incidencia en grupos etarios específicos, etc.). La *green zone strategy* o estrategia de supresión se ha implementado en países-isla o con posibilidad de convertirse en territorios estancos (Nueva Zelanda y Australia, China y Estados-isla como Taiwán). Europa adoptó la estrategia de mitigación (*red zone strategy*). La mayor parte de la literatura económica aboga por esta última, basada en el cierre imprescindible, y destaca el enorme valor de los test individuales para mantener el tejido productivo en actividad. La literatura de salud pública, en cambio, aboga más por la estrategia de supresión (Oliu-Barton y Pradelski, 2021). Hoy por hoy no hay estudios rigurosos que permitan establecer la dominancia de una estrategia sobre la otra, y los escasos estudios (Zala *et al.*, 2020) se basan en proyecciones de modelos epidemiológicos que ya han demostrado ser erróneas.

Idealmente, serían necesarios modelos epi-económicos capaces de predecir las consecuencias

sobre la salud y sobre la economía de cursos de acción alternativos. Los *outcomes* son de salud (fundamentalmente, fallecimientos) y logros económicos. Como fuente de variación exógena, algunos modelos incorporan datos del pasado, sanitarios (incidencias, hospitalizaciones) y no sanitarios (intensidad de las restricciones entre otros) en distintas zonas o países, con el fin de anticipar los efectos, sobre la salud y económicos, de los posibles cursos de acción en la gestión macro de la pandemia. En España, las medidas más restrictivas fueron las de confinamiento domiciliario entre marzo y mayo/junio de 2020. Según una evaluación basada en modelos econométricos espaciales con datos provinciales (Orea y Álvarez, 2022), el confinamiento pudo haber reducido el número de casos en un 82,8 por 100. De haber confinado una semana antes, se habrían podido evitar un 11,6 por 100 adicional de casos y se habría conseguido un ahorro económico muy considerable.

La evaluación cuantitativa, aunque importante, no constituye el único método de evaluación útil, particularmente cuando lo que debe evaluarse es un proceso complejo de respuesta a una crisis de salud pública como la derivada de la epidemia de COVID-19, que ha tenido consecuencias en múltiples dimensiones. Ya a mediados de 2020, la OMS puso en marcha un proceso de evaluación independiente y exhaustivo con el fin de valorar la respuesta sanitaria internacional a la pandemia, del que pueden extraerse recomendaciones importantes para mejorar las capacidades de respuesta futuras (Sirleaf y Clark, 2021).

Sobre la base de la guía de la OMS para las denominadas «after-action reviews» (AAR) o «revisiones posactuación» (World Health Organization, 2019), el ECDC ha proporcionado un detallado protocolo de evaluación de las decisiones de respuesta a la pandemia. De acuerdo con el mismo, las AAR buscan evaluar las acciones de respuesta a un evento preocupante para la salud pública mediante la observación objetiva, el análisis de problemas y/o mejores prácticas, y la identificación de áreas de mejora en las actividades de preparación y respuesta, que suma a la recopilación de la evidencia científica disponible un proceso de reflexión que involucra a todos los agentes relevantes (ECDC, 2021).

Algunos países han publicado ya informes de evaluación que, sin seguir estrictamente el protocolo de las AAR, constituyen ejemplos de evaluación cualitativa orientada a identificar las lecciones que deben aprenderse de la gestión de la pandemia. Es el caso de Inglaterra, que explicita en su informe su espíritu constructivo, alejado de la búsqueda de culpables, un elemento que ha de considerarse crucial para el éxito y credibilidad de la evaluación (House of Commons, 2021). En España se ha reclamado una evaluación global de la gestión de la pandemia desde las primeras oleadas (García-Basteiro *et al.*, 2020; Martín-Moreno *et al.*, 2020) por el gran impacto en mortalidad, particularmente en residencias. El Consejo Interterritorial del Sistema Nacional de Salud (SNS) ha hecho el encargo en septiembre de 2021, y en el momento de publicación de este artículo el informe se encuentra en proceso de elaboración.

2. El principio del coste-efectividad como telón de fondo. Evaluación de las decisiones de cobertura, financiación de prestaciones del SNS y de la transformación digital

En el ámbito de la política sanitaria, la evaluación por excelencia ha sido la que consiste en aplicar el análisis coste-efectividad a los tratamientos, tecnologías e intervenciones sanitarias. La evaluación económica de procedimientos se ha estandarizado metodológicamente en el mundo, e incluso se ha convertido en una profesión, la farmacoeconomía, por ser los medicamentos el objeto más habitual de análisis. En cuanto a la metodología empleada, se vela por el uso de protocolos que impidan influencias indebidas o interesadas en los resultados, por ejemplo, a través del registro previo de los proyectos de evaluación, y se van produciendo avances importantes en lo que respecta a distintos aspectos, como por ejemplo el alcance de los costes indirectos.

Más allá del marco estrecho de los tratamientos médicos individuales, en sanidad hay que evaluar también cambios organizativos, intervenciones preventivas a nivel comunitario y políticas que inducen efectos fuera del sector sanitario y de la salud. Hay métodos como el *SROI* (*social return of investment*) que se han diseñado para ello y son de obligada aplicación para intervenciones sociales o, por ejemplo, para justificar actividades de las ONG en algunos países (para una revisión de aplicaciones del *SROI* en sanidad, véase González López-Valcarcel, 2019). Estos métodos se aplican cada vez más para evaluar políticas públicas

multisectoriales y del tercer sector. Aunque se ha avanzado en estandarización metodológica, todavía hay un gran trecho por recorrer.

La institucionalización de la evaluación económica de tecnologías sanitarias en España

Los desarrollos teórico-metodológicos y una enorme cantidad de aplicaciones de análisis de coste-efectividad en sanidad, accesibles en bases de datos como el registro de la Universidad de Tufts (<https://cevr.tuftsmedicalcenter.org/databases/cea-registry>) o el *Centre for Reviews and Dissemination* de la Universidad de York (<https://www.york.ac.uk/crd/>), han sido de gran utilidad en entornos institucionales en los que una cuarta barrera (demostrar el coste-efectividad) es legalmente preceptiva para incorporar una nueva tecnología a la cartera de prestaciones públicas. La mayor parte de países de la UE-15 exigen esa cuarta barrera actualmente.

España es una excepción, pues todavía no se reconoce formalmente la cuarta barrera para que un nuevo medicamento o prestación entre en la cartera de servicios y sea financiado por el SNS. A lo largo del tiempo, la necesidad de evaluación económica y el reconocimiento del criterio de coste-efectividad para la financiación pública de prestaciones se han ido incorporando de forma oblicua en el marco normativo (cuadro n.º 1).

La institucionalización de la evaluación de tecnologías sanitarias en España ha avanzado en los últimos años, pero todavía no está completada. Desde 2006 se ha ido conformando una red de colaboración entre agencias

CUADRO N.º 1

EL COSTE-EFECTIVIDAD Y LA EVALUACIÓN ECONÓMICA COMO CRITERIOS EN ESPAÑA. MARCO REGULATORIO

NORMA	CONTENIDO
Real Decreto 63/1995 sobre ordenación de prestaciones sanitarias.	«Para incorporar nuevas técnicas, tecnologías y procedimientos, la administración sanitaria del Estado ha de valorar su seguridad, eficacia y eficiencia».
Real Decreto 1030/2006 por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización.	Para actualizar la cartera de servicios comunes del SNS «se deberá utilizar el procedimiento de evaluación más adecuado en cada caso que permita conocer el coste, la eficacia, la eficiencia, la efectividad, la seguridad y la utilidad sanitaria».
Real Decreto-ley 16/2012, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud.	La inclusión de medicamentos en la financiación del Sistema Nacional de Salud tendrá en cuenta (entre otros) el criterio del «valor terapéutico y social del medicamento y beneficio clínico incremental del mismo teniendo en cuenta su relación coste-efectividad». Las nuevas técnicas, tecnologías o procedimientos serán sometidas a evaluación, con carácter preceptivo y previo a su utilización en el Sistema Nacional de Salud, por la RedETS. «Para la decisión de financiación de nuevos medicamentos, además del correspondiente análisis coste-efectividad y de impacto presupuestario, se tendrá en cuenta el componente de innovación, para avances terapéuticos indiscutibles (...) y su contribución a la sostenibilidad del Sistema Nacional de Salud». «La Comisión Interministerial de Precios de los Medicamentos tendrá en consideración los análisis coste-efectividad y de impacto presupuestario».
Orden SSI/1833/2013, por la que se crea y regula el Consejo de la RedETS.	«La Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud tiene como misión generar, difundir y facilitar la implementación de información destinada a fundamentar la toma de decisiones en el Sistema Nacional de Salud, contribuyendo de esta forma al incremento de la calidad, equidad, eficiencia y cohesión en el mismo». El pleno aprobará «el Plan Anual de Evaluación de Tecnologías Sanitarias (...) que se someterá a informe del Pleno del Consejo Interterritorial del Sistema Nacional de Salud».
Real Decreto Legislativo 1/2015 por el que se aprueba el texto refundido de la Ley de garantías y uso racional de los medicamentos y productos sanitarios.	«La ficha técnica o resumen de las características del producto (...) A la ficha técnica se acompañará, preceptivamente, información actualizada del precio del medicamento y, cuando sea posible, la estimación del coste del tratamiento» (art.15).

Fuente: Elaboración propia.

de evaluación de tecnologías sanitarias, algunas de las cuales ya existían desde principios de los años noventa. La red se constituye formalmente por el Consejo Interterritorial del SNS en 2012 como Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) (<https://redets.sanidad.gob.es/>). RedETS «pretende fundamentar la toma de decisiones sobre incorporación, condiciones de financiación/desinversión, y uso apropiado de las tecnologías sanitarias, con el fin de promover la equidad y sostenibilidad del SNS» (3). Actualmente está formada por ocho unidades, siete de comunidades autónomas (Andalucía, Aragón,

Canarias, Cataluña, Galicia, Madrid, País Vasco) y el Instituto de Salud Carlos III, que colaboran en red, trabajan por encargo y con financiación del Ministerio de Sanidad y de los Gobiernos regionales, y elaboran informes no preceptivos sobre determinadas tecnologías, destacando las no farmacológicas (dispositivos médicos, procedimientos, etc.). Aunque explícitamente no hay una especialización «oficial» de las agencias, en la práctica algunas trabajan más intensamente determinados temas. Además de la función *primaria* de evaluación de tecnologías sanitarias (1 y 2 en la lista siguiente), la mayor parte de unidades de la red asumen otras funciones:

1. Informes de evaluación de tecnologías sanitarias (procedimientos, dispositivos, pruebas diagnósticas...).
2. Informes de evaluación de medicamentos (estudios coste-efectividad).
3. Guías de práctica clínica.
4. Proyectos de investigación (convocatorias competitivas nacionales e internacionales).
5. Compra pública innovadora.
6. Radar de detección de tecnologías emergentes a evaluar.
7. Guía de desinversión (detección de tecnologías obsoletas o que no aportan valor).

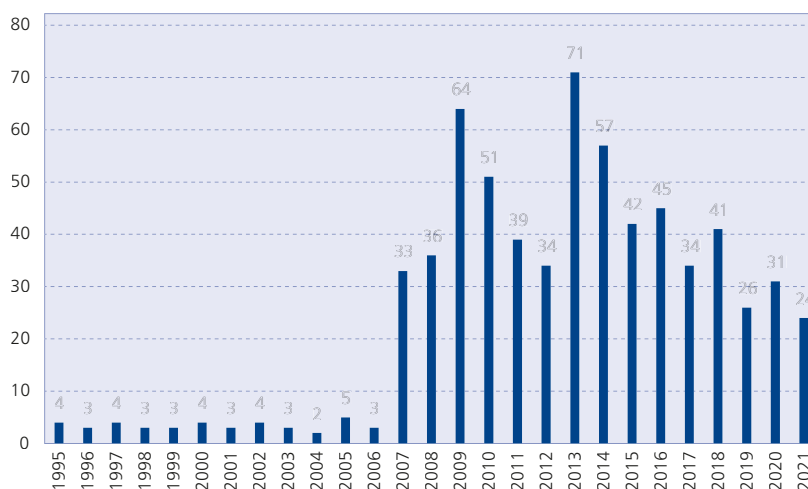
8. Metodología de la evaluación.
9. Formación y transferencia de conocimiento.

Algunas de estas agencias forman parte de unidades gubernamentales más amplias de investigación o generación y transferencia de conocimiento. Aunque se suele hablar de la «red de agencias», jurídicamente han ido evolucionando hacia otras formas jurídicas. En la mayor parte de los casos son unidades administrativas en el organigrama de las consejerías de sanidad (Andalucía, Madrid) o del servicio regional de salud (Canarias). AQuAS, por su parte, sí conserva el estatus de agencia independiente.

La RedETS aporta valor al SNS y aprovecha economías de alcance y de red. A su vez, las organizaciones de la RedETS participan en la Red Europea de Agencias de Evaluación (EUnetHTA). El nivel de recursos, actividad y producción de informes es muy desigual entre los miembros de la red. La más prolífica es, sin duda, la agencia catalana. En 2019, cuando AQuAS cumplía 25 años, había elaborado 1.130 informes de evaluación y publicado 750 artículos en revistas científicas (4). Durante las últimas dos décadas ha habido un enorme aumento tanto de la actividad de evaluación de tecnologías sanitarias en España como del número de miembros de la red. Considerando únicamente los informes de evaluación de tecnologías sanitarias en sentido estricto, hasta 2022 el conjunto de los ocho miembros actuales de la RedETS había publicado 669 informes de evaluación de tecnologías sanitarias (5) (gráfico 1). Hasta el año 2000, se publicaban una media de 3,5 informes anuales; en la

GRÁFICO 1

NÚMERO ANUAL DE INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS EN ESPAÑA (RedETS) 1995 A 2021



Fuente: RedETS (2022). Listado de informes de evaluación de tecnologías sanitarias. Accesible en <https://redets.sanidad.gob.es/productos/buscarProductos.do?metodo=buscaTipos&tipoid=1>

década 2001-2010, 20,4; y en la última década 40,4 informes anuales. El 30 por 100 de los informes son posteriores a 2015.

Pese al indudable avance en el conocimiento, su aplicación sistemática a la toma de decisiones sigue postergada en nuestro país. En 2021, un grupo de académicos y profesionales de la sanidad y de la economía en España reivindicaba la constitución de una agencia independiente de evaluación de prácticas y políticas sanitarias, al estilo de la National Institute for Health and Care Excellence (NICE) británica, con capacidad legal para recomendar e inducir a los Gobiernos a cumplir las recomendaciones o, alternativamente, a explicar por qué no las cumplen. La iniciativa llegó a la comisión de sanidad del Congreso de los Diputados en noviembre de 2021 con una proposición no de ley (6), pero no pasó de ahí.

Hasta ahora, los pasos dados en este ámbito han sido claramente insuficientes. El Ministerio de Sanidad ha reconocido y asumido la metodología de evaluación de medicamentos del grupo GENESIS, una red de servicios de farmacia hospitalaria, como «oficial» para elaborar los informes de posicionamiento terapéutico (IPT), cuyo objetivo es analizar el valor añadido de los nuevos medicamentos. La ventaja de los informes GENESIS es su disponibilidad inmediata en tiempo real para tomar decisiones de compra, pero se trata de evaluaciones exprés que, a pesar de seguir una metodología común y consensuada, no siguen los estándares internacionales y están lejos de ser evaluaciones completas, lo que requiere de mucho tiempo y recursos. Los IPT sobre nuevos medicamentos se elaboran y publican desde 2013 por el Ministerio de Sanidad, pero hasta 2021 únicamente consideraban eficacia y efectividad,

sin incorporar información ni consideraciones de costes ni de coste-efectividad.

La preocupación específica por los medicamentos de alto coste también ha llevado a la creación de VALTERMED (Sistema de Información de Valor Terapéutico de Medicamentos), un sistema de información para determinar el valor terapéutico en la práctica clínica real de los medicamentos de alto impacto sanitario y económico en el SNS (<https://www.mscbs.gob.es/profesionales/farmacia/valtermed/home.htm>). Es una base de datos creada en noviembre de 2019, de acceso restringido. En diciembre de 2021, con dos años de existencia, había 13 protocolos fármaco-clínicos publicados en la web. Los informes podrían afectar a más de once mil pacientes (7).

Los mayores propulsores del gasto sanitario, particularmente los medicamentos oncológicos y otros de uso hospitalario, así como los tratamientos personalizados (CAR-T) constituyen, por su alto coste e impacto presupuestario, una prioridad en la evaluación. El disponer de criterios sociales de disposición a pagar por un año de vida ajustado por calidad, con posibilidad de diferenciación por grupos de pacientes o patologías, es una de las cuestiones abiertas. En el Reino Unido, que ha sido un pionero en este ámbito, la agencia de evaluación NICE está en proceso de redefinir umbrales por patologías.

En resumen, la arquitectura institucional de la evaluación de tecnologías sanitarias y medicamentos en España es compleja y poco resolutive, no hay cuarta barrera ni se ha declarado umbral de Ratio de Coste Efectividad

Incremental, tampoco existen criterios explícitos de priorización de grupos de pacientes o enfermedades, aunque se reconoce la necesidad de priorizar. Comparándonos con otros países europeos, «aún falta garantizar la independencia de las agencias de evaluación de tecnologías sanitarias, eliminar la influencia de la política económica e industrial de los comités de decisiones sanitarias, fijar precios basados en el valor terapéutico, mejorar la transparencia del proceso y de los resultados de las evaluaciones y, por último, aumentar la participación de los grupos de interés. Al igual que en otros países de su entorno, España tiene que clarificar el papel de la evaluación económica como criterio de decisión» (Epstein y Espín, 2020).

El uso de la evaluación económica (coste-efectividad) ligado a la pandemia: rastreo y vacunas

Hay dos tecnologías de particular interés relacionadas con la pandemia: el rastreo de contactos (que forma parte de una estrategia más general de test, rastreo y aislamiento [TRA]), y las vacunas.

En la medida en que la situación epidemiológica y el conocimiento sobre la enfermedad han ido cambiando, los análisis de coste-efectividad realizados para unas condiciones locales determinadas han ido perdiendo validez externa. A la altura de noviembre de 2020, estimábamos que la ratio beneficio/coste de la estrategia TRA en España podría estar en torno a 7/1: por cada euro gastado se recuperarían 7 en el medio plazo (Vallejo Torres y González López-Valcárcel, 2021). Por tanto, el TRA es una estrategia dominante en el contexto de una

determinada transmisibilidad e incidencia. Ese resultado depende de la incidencia acumulada y de la relación entre incidencia y uso de recursos sanitarios, y entre incidencia y fallecimientos. Dichas relaciones han cambiado sustancialmente con las variantes que han ido surgiendo, especialmente con ómicron. La clara dominancia de la estrategia TRA se debe, fundamentalmente, al hecho de que los principales costes de la COVID son costes indirectos, generados a través de los efectos en el tejido productivo y la economía.

Lo mismo ocurrió con las vacunas. Un estudio publicado en *Science* (Castillo *et al.*, 2021) por equipos de científicos de primera línea estimaba el coste-efectividad de expandir la capacidad de producción de vacunas en el mundo, acortando al máximo los tiempos hasta la inmunidad de grupo y la vuelta a la normalidad económica en los países de renta baja y media. Los resultados señalaban que el valor marginal de una dosis de vacuna en 2021 si se consiguiera expandir en mil millones de dosis la capacidad prevista de producción para ese año podría estar en torno a los 600-1.000 dólares. Un análisis estándar de coste-efectividad recomendaría comprar hasta ese precio, lo que resulta obviamente disparatado.

Otra estimación para la región metropolitana del norte de Cataluña (López *et al.*, 2022) basada en un análisis coste-beneficio concluía que la vacunación tiene una ratio beneficio/coste de 3,4 desde la perspectiva social (1,4 desde la perspectiva del sistema sanitario), con unos beneficios sociales de 116,67 euros por dosis de vacuna. La enorme disparidad entre estos resultados

y los del artículo de Castillo y colaboradores (Castillo *et al.*, 2021) radica en que este último considera también los costes y ahorros en el sistema económico que se producen si las vacunas ganan tiempo en la vuelta a la normalidad. El estudio para Cataluña no considera los ahorros derivados de evitar cierres por una vuelta adelantada a la actividad económica normal, que es la principal fuente de ahorro de costes en el estudio de Castillo. Como ya hemos señalado, los mayores costes inducidos por la COVID son sobre la economía.

Por otra parte, la incertidumbre ha marcado la validez de los análisis. La ratio de coste-efectividad incremental (RCEI) de las vacunas, a nivel mundial, se estimaba tan alta en la primavera de 2021 porque en aquel momento se confiaba o suponía que la vacuna inmunizaba durante un plazo largo, sin necesidad de dosis de refuerzo, y que la inmunidad de grupo se alcanzaría con un (indeterminado) alto porcentaje de vacunados. Unos meses más tarde se sabía que ninguno de esos dos supuestos era realista. En consecuencia, el coste-efectividad de las vacunas será muy inferior al estimado inicialmente.

3. Pero no todo es COVID. Evaluaciones de políticas públicas recientes en sanidad/salud y apuntes para una agenda priorizada

La evaluación en sanidad debe abarcar los tres niveles: macro (las políticas de salud y sanitarias), meso (la gestión y organización de centros y servicios sanitarios) y micro (la gestión clínica). Una parte importante

de la evaluación es la de cambios organizativos. España, por su descentralización sanitaria, cuenta con una enorme cantidad de experiencias que, si se evalúan bien, permitirán la mejora por comparación, el *benchmarking* y el escalado desde experiencias piloto hacia la generalización en todo el territorio. AQUAS aporta un valioso repositorio de experiencias de innovación en gestión y organización del que se puede aprender. Otra iniciativa interesante viene dada por los Premios Profesor Barea de Gestión Sanitaria, con los que desde 2002 la Fundación Signo viene reconociendo a centros y grupos de profesionales que llevan a cabo innovaciones valiosas (y evaluables) de gestión sanitaria (8).

En cuanto a políticas macro, la agenda de evaluación de intervenciones sanitarias de la Autoridad Independiente de Responsabilidad Fiscal española (AIReF) constituye un buen punto de partida. Refleja las prioridades del Gobierno, impelidas por las presiones del gasto sanitario. Los *spending review* (9) I y II han evaluado el gasto en medicamentos de prescripción, medicamentos de uso hospitalario y las inversiones sanitarias en bienes de equipo. Los resultados de dichas evaluaciones sugieren líneas de avance.

Sin pretensión de exhaustividad, se proponen aquí algunos ejemplos de evaluación en el área sanitaria, que apuntan hacia grandes áreas en las que se ha intervenido o es necesario hacerlo. Se han seleccionado casos de macrogestión, de mesogestión y de evaluación de los efectos sobre la salud de determinadas políticas económicas y fiscales.

La *digitalización* de la sanidad puede ser un legado posi-

tivo a largo plazo de la COVID-19. La pandemia ha acelerado un proceso en marcha, en algunos casos de forma brusca. España ha sido el país de la UE con mayor uso de consultas remotas en atención primaria (en marzo de 2021, más del 70 por 100 de la población había tenido alguna, frente al 38 por 100 del conjunto de la UE). Es el país de la Unión con mayor tasa de certificados COVID emitidos. La transformación digital de la sanidad impone nuevos retos para la evaluación económica de tecnologías sanitarias (González López-Valcárcel, Torres y Domínguez, 2021). La propuesta de la NICE, que diferencia entre tecnologías digitales según su individualización en pacientes concretos y el grado de intromisión en su vida, es un buen punto de partida para reformular estándares de evaluación económica. La UE señala la falta de evidencia sobre lo que funciona y lo que no funciona en la sanidad digital, y la necesidad de mirar más allá de la emergencia pandémica para evaluar cómo deberían reajustarse las tecnologías digitales, evitando las consecuencias adversas no intencionadas de su uso, y también cómo incentivar su uso (European Commission, DG SANTE, 2021). La evaluación sistemática de la digitalización en el sector sanitario y de las tecnologías digitales y el diseño y adaptación de la metodología de evaluación son tareas pendientes.

Evaluación de la gestión indirecta de la sanidad, en sus distintas formas y modalidades y de forma más general de las grandes políticas de concertación y colaboración público-privada en la gestión sanitaria. El llamado «modelo Alzira», de contratación externa de toda la gestión del

aseguramiento público y prestaciones de servicios con una entidad externa se puso en marcha en 1999 en la Comunidad Valenciana, primero solo para atención hospitalaria, ampliándose a la atención primaria en 2003. El modelo se extendió por toda la Comunidad, donde en torno al 20 por 100 de la población vivía en áreas de salud cubiertas por este tipo de acuerdos. En 2018, fecha de prescripción del contrato de quince años con la primera prestadora (Ribera Salud) comenzó el proceso de reversión para la comarca de Alzira. A pesar de que el modelo llevaba en funcionamiento casi veinte años y que llegó a convertirse en un escaparate de la modernidad en programas MBA de prestigio, como estudio de caso (10), no hay publicaciones científicas con estudios rigurosos de evaluación comparativa de la calidad y eficiencia de gestión entre los dos modelos. Un informe de 2017 encargado por la concesionaria (López-Casasnovas y del Llano, 2017) comparó los hospitales valencianos de la concesionaria, ajustando por complejidad y casuística, con hospitales catalanes, que formaron el comparador sintético. No se pudo hacer la comparación directa con los hospitales públicos de la Comunidad Valenciana por falta de información. Concluía que no hay diferencias significativas ni en indicadores clínicos ni económicos con los hospitales catalanes comparables. El propio informe advertía que la evaluación presenta limitaciones, fundamentalmente debidas a la falta de datos objetivos para poder establecer un comparador sintético creíble.

También se ha generado evidencia a través de análisis de frontera eficiente, que comparan

la productividad y calidad de hospitales españoles. La mayor parte de estudios sugieren que no es la forma jurídica ni la propiedad (hospital público o privado), sino la capacidad de gestionar recursos, la holgura de gestión, la que se asocia significativamente con mejores resultados (Comendeiro-Maaløe *et al.*, 2019). Una excepción es el estudio de Ortega y colaboradores (Ortega-Díaz *et al.*, 2020), que encuentra mayores probabilidades de estar en la frontera eficiente a los centros en régimen de colaboración público-privada en el período 2002-2012.

Evaluación de los efectos del cambio de copago de medicamentos de prescripción sobre la eficiencia, la equidad y la adherencia de los pacientes. En 2012 cambió el copago de medicamentos de prescripción en España. Los pensionistas, que tenían medicamentos gratuitos, empezaron a pagar el 10 por 100 o más del precio, con límites mensuales según renta. Los activos siguieron pagando, sin límites, el 40 por 100 del precio o subieron al 50 por 100 o 60 por 100 según renta, en función de tres tramos (base liquidable en IRPF hasta 18.000 euros anuales, entre 18.000 y 100.000 euros y más de 100.000 euros). Las evaluaciones publicadas, con distintos datos y métodos de análisis, apuntan a que la elasticidad-precio de la demanda de medicamentos es muy baja y en consecuencia el principal efecto ha sido el de redistribuir la carga financiera entre el Gobierno y las familias, si bien a corto plazo ha habido reducciones importantes del consumo de medicamentos, también de los esenciales como los que se han de tomar después de sufrir un infarto de miocardio (González López-Valcárcel *et al.*, 2017). Los efectos sobre la equidad han

sido mucho mayores que sobre la eficiencia y eliminar la diferenciación, artificial y trasnochada, entre «activos» y «pensionistas» se configura como una vía de avance técnicamente adecuada, pero políticamente costosa.

Evaluación de intervenciones económicas y fiscales sobre la salud de la población. Es el caso de políticas de salud pública basadas en medidas económicas y fiscales, como los impuestos a las bebidas azucaradas. El impuesto catalán de 2017 ha tenido impacto significativo de reducción de ventas (Vall Castelló y López Casasnovas, 2020) con efectos heterogéneos entre tipos de familias (Fichera *et al.*, 2021).

También se ha evaluado el Régimen Específico de Abastecimiento de Canarias (REA) desde la perspectiva de la salud. Las subvenciones a ciertos alimentos importados para garantizar el aprovisionamiento local de las islas han resultado en un daño colateral para la salud. Si se suprimieran las subvenciones al azúcar, mantequilla y queso se podrían evitar más de diez mil casos de obesidad en adultos en Canarias (Rodríguez-Mireles *et al.*, 2021).

Seguridad del paciente e intervenciones sanitarias de bajo valor. Una parte importante de intervenciones sanitarias son de bajo valor terapéutico, es decir, de baja efectividad. Detectar cuál es su prevalencia por centros y zonas de salud es un primer paso para avanzar hacia el paradigma de un sistema sanitario basado en el valor. En España, afortunadamente, se cuenta con información transparente y accesible de la frecuencia de uso de procedimientos de bajo valor (cesáreas

en embarazos de bajo riesgo, tonsilectomía en niños, liberación del túnel carpiano, etc.), gracias al *Atlas de variaciones en la práctica médica* (<https://www.atlasvpm.org/>), que también analiza las variaciones injustificadas en la calidad de los cuidados hospitalarios en España. La efectividad de muchos procedimientos y tratamientos médicos tiene gran variabilidad entre pacientes, de modo que una misma intervención puede ser muy efectiva en uno y muy poco en otro. Esas tecnologías «tipo II», en la terminología de Chandra y Skinner (Chandra y Skinner, 2012), son las que hay que protocolizar con más cuidado, y deben monitorizarse sus indicaciones para evitar sobreutilización que aporta escaso valor. No hay ningún riesgo, en cambio, de utilización inapropiada de un antiviral contra el VIH, que es efectivo en todos los pacientes.

En la evaluación en sanidad es importante cuantificar no solo el coste de lo que se hace, y su relación con el beneficio terapéutico, sino también los costes que se podrían evitar con una práctica correcta. Las hospitalizaciones potencialmente evitables aproximan de alguna manera la calidad del sistema de salud. Hay diferencias en Europa, como ha puesto en evidencia un estudio comparativo de seis países basado en microdatos homogéneos de pacientes y para varias enfermedades crónicas (Thygesen *et al.*, 2015). De forma paralela, es importante mejorar la medición de los resultados (salud): por ejemplo, es mejor hacerlo con la mortalidad prematura y sanitariamente evitable (MIPSE) que con la esperanza de vida. La MIPSE recoge las muertes producidas por causas que el sistema sanitario es capaz de evitar (posponer).

Trabajar en la seguridad del paciente y limitar la medicina defensiva (pedir pruebas o realizar procedimientos innecesarios para evitar posibles demandas) con incentivos y regulación adecuados son dos grandes retos de los sistemas sanitarios en los países desarrollados, que necesitan urgentemente evaluación *ex post* y *ex ante*. La OCDE estima que el daño a los pacientes que infringe el propio sistema sanitario representa en torno al 12,6 por 100 de todo el gasto sanitario, y más de la mitad (8,7 por 100) es prevenible (OCDE, 2020).

4. Evaluaciones pendientes necesarias para tomar decisiones. Apuntes para una agenda de priorización

En su informe 2021 sobre la salud en Europa, la Comisión Europea (European Commission, DG SANTE, 2021) señala tres prioridades: 1) conocer las consecuencias de la pandemia sobre la salud de las poblaciones, la COVID persistente y los efectos colaterales en la atención de otras enfermedades y su abordaje; 2) implementar una transformación digital de la sanidad con una hoja de ruta de evaluación para priorizar y redirigir; y 3) atender a la necesidad de más recursos humanos para la salud, más flexibles y con nuevos perfiles profesionales.

La evaluación de los efectos de la COVID sobre la salud en el medio plazo, como consecuencia de los retrasos en la atención, es un tema imperativo, que conecta directamente con el de la priorización y gestión de listas de espera. Entre los efectos a largo plazo de la COVID, los problemas de salud mental y las respuestas del sistema sanitario para

abordarlos requieren evaluación prospectiva.

Las desigualdades en el uso de servicios sanitarios, que pueden derivar en desigualdades en salud, han de ser evaluadas también. En este sentido, apuntamos la necesidad de estudiar cambios de diseño de los copagos que incorporen criterios de eficiencia, equidad y objetivos presupuestarios.

En cuanto a la fiscalidad del sector sanitario y la financiación de las prestaciones, convendría tener evaluaciones prospectivas de posibles cambios impositivos en el sector como, por ejemplo, la eliminación de las exenciones en IRPF/sociedades por los seguros de salud, y debería revisarse, como ya se ha apuntado, la arquitectura institucional de la evaluación económica de medicamentos y otras tecnologías para decidir cobertura y precio, garantizando la transparencia y la rendición de cuentas de las decisiones tomadas.

El llamado modelo MUFACE (Mutualidad General de Funcionarios Civiles del Estado) es una excepción española perdurable. Casi dos millones de funcionarios civiles del Estado, de justicia y de las fuerzas armadas tienen la posibilidad de elegir aseguradora entre las conveniadas, tanto privadas como el propio Instituto Nacional de la Seguridad Social. Ese modelo se ha puesto como ejemplo a seguir e incluso generalizar, desde posiciones de parte. Es una caja negra en la sanidad española, los datos que permitan comparar las prestaciones, acceso y salud lograda para los asegurados de ambas redes no están disponibles. El sesgo de autoselección lleva a que los *pool* de riesgos de ambos

colectivos no sean comparables, ni por tanto el coste unitario medio por persona. Una evaluación objetiva sería necesaria para fundamentar los debates y tomar decisiones.

Por último, un ensayo de políticas de gran interés para la evaluación es el área sanitaria única de Madrid, y convendría medir sus efectos sobre la eficiencia y la calidad de las prestaciones sanitarias públicas, así como los efectos cruzados sobre la utilización y financiación de los centros propios de la red pública madrileña. Con los datos disponibles del Observatorio Madrileño se podría hacer una primera aproximación.

III. CONCLUSIONES

1. En sanidad, las evaluaciones *ex post* de las intervenciones y políticas públicas son útiles, pero cada vez resulta más necesaria la evaluación prospectiva *ex ante* para facilitar la toma de decisiones fundamentada en datos.

2. Ha habido un enorme avance del conocimiento durante los años de pandemia 2020-2021, en parte gracias al uso de datos masivos compartidos, de acceso abierto y en tiempo real. La inteligencia artificial, que hace uso de esos datos, es complementaria, no sustitutiva, de la inteligencia humana y del conocimiento de los problemas cuyos datos se analizan.

3. Los métodos de evaluación de las políticas públicas se apoyan cada vez más en experimentos sociales y en la disponibilidad de registros y de datos cuasiexperimentales, así como en la posibilidad de explotar grandes masas de datos, datos de

vida real, estructurados o no, y de realizar megaensayos.

4. Los modelos epieconómicos son de enorme utilidad y pueden ayudar a la toma de decisiones, pero sirven para lo que sirven: explicar el pasado, predecir a corto plazo, retratar el presente (*nowcasting*) y avanzar hipótesis plausibles sobre el futuro (con enorme incertidumbre). Igual que la gestión de la pandemia, los modelos deben ser evaluados retrospectivamente, según su capacidad predictiva.

5. El mayor reto de los países durante estos dos años de pandemia ha consistido en encontrar soluciones para compatibilizar la pandemia con la vida de las personas y mantener el tejido productivo y el crecimiento económico a medio plazo. Nunca la sanidad y la economía habían sido tan interdependientes, nunca las políticas sanitarias habían tenido tanto de políticas económicas.

6. A grandes rasgos, la gestión de la pandemia abarca: 1) las medidas para limitar los contagios; 2) la disposición y organización de recursos y servicios sanitarios y sociosanitarios para prevenir y tratar la enfermedad; 3) las medidas de carácter económico para minimizar los efectos de la pandemia sobre el tejido productivo y sobre las economías familiares.

7. Hoy por hoy no hay estudios rigurosos que permitan establecer la dominancia de la estrategia de supresión sobre la de contención de la COVID, o viceversa.

8. Los métodos de investigación cualitativa deben reivindicar su utilidad, particularmente en la evaluación de la respuesta a

la pandemia. Las denominadas «after-action reviews» (AAR) o «revisiones posactuación» pueden servir para identificar las lecciones aprendidas y diseñar respuestas futuras más eficaces de forma constructiva, sin necesidad de identificar culpables.

9. La institucionalización de la evaluación de tecnologías sanitarias en España ha avanzado en los últimos años, pero todavía no está completada. La arquitectura institucional de la evaluación de tecnologías sanitarias y medicamentos es compleja y poco resolutive, y no existen criterios explícitos de priorización de grupos de pacientes o enfermedades, aunque se reconoce la necesidad de priorizar y se acepta el criterio de coste-efectividad.

10. Durante la pandemia quedó en suspenso el coste-efectividad como criterio de decisión, tanto por imperativo moral como por la apabullante evidencia de que prácticamente todo gasto sanitario en prevención y tratamientos COVID es dominante, teniendo en cuenta los costes (y ahorros de costes) indirectos por la vía del impacto en la economía.

11. La evaluación en sanidad debe abarcar los tres niveles: macro (las políticas de salud y sanitarias), meso (la gestión y organización de centros y servicios sanitarios) y micro (la gestión clínica). Una parte importante de la evaluación es la de cambios organizativos. España, por su descentralización sanitaria, cuenta con una enorme cantidad de experiencias que, si se evalúan bien, permitirán la mejora por comparación, el *benchmarking* y el escalado desde experiencias piloto hacia la generalización en todo el territorio.

12. En cuanto a políticas macro, la agenda de evaluación de intervenciones sanitarias de la AIReF constituye un buen punto de partida. Los resultados de dichas evaluaciones sugieren líneas de avance a considerar.

13. Hay áreas de especial relevancia para la evaluación: la digitalización de la sanidad, la gestión indirecta de la prestación sanitaria, las tecnologías de alto coste unitario (medicina personalizada entre otras). Hay que evaluar las intervenciones económicas y fiscales sobre la salud de la población, y analizar la actividad sanitaria con perspectiva de sistema basado en el valor. Los mayores propulsores del gasto sanitario, particularmente los medicamentos oncológicos y otros de uso hospitalario, así como los tratamientos personalizados (CAR-T) constituyen, por su alto coste e impacto presupuestario, una prioridad en la evaluación. El disponer de criterios sociales de disposición a pagar por un año de vida ajustado por calidad, con posibilidad de diferenciación por grupos de pacientes o patologías, es una de las cuestiones abiertas.

14. Finalmente, el modelo MUFACE requiere una evaluación urgente. En la agenda inmediata deberían entrar también las políticas sanitarias y de salud para hacer frente a las consecuencias a largo plazo de la COVID sobre la salud (también mental) de la población y la organización de los recursos necesarios para abordar el problema, así como su financiación. Las desigualdades en el uso de servicios sanitarios, que pueden derivar en desigualdades en salud, han de ser evaluadas sin demora.

NOTAS

(*) Agradezco a ROSA URBANOS los comentarios y aportaciones al texto.

(1) Ver GONZÁLEZ LÓPEZ-VALCÁRCEL *et al.* (2020). <https://theconversation.com/covid-19-pandemia-de-modelos-matematicos-136212>

(2) <https://covid19forecasthub.eu/>

(3) Presentación de la Agencia de Evaluación de Tecnologías Sanitarias del ISCIII. <https://www.isciii.es/QuienesSomos/CentrosPropios/AETS/Paginas/default.aspx>

(4) https://aquas.gencat.cat/ca/sobre_aquas/presentacio/commemoracio-25-anys-agencia

(5) Una base de datos en Excel con el listado actualizado de los informes está disponible en <https://redets.sanidad.gob.es/productos/buscarProductos.do?metodo=buscaTipos&tipoid=1>

(6) <https://www.diariofarmacia.com/2021/11/21/el-pdecat-lleva-al-congreso-el-debate-sobre-impulsar-el-hispanice>

(7) <https://www.redaccionmedica.com/secciones/industria/los-protocolos-de-remdesivir-y-dupilumab-copan-las-visitas-de-valter-med-4905>

(8) <https://www.fundacionsigno.com/premios.php>

(9) <https://www.airef.es/es/spending-review/>

(10) <https://www.eleconomista.es/valencia-na/noticias/8673190/10/17/El-modelo-de-gestion-sanitaria-de-Ribera-Salud-caso-de-estudio-en-Harvard.html>

BIBLIOGRAFÍA

ANGRIST, J. D. y PISCHKE, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2), pp. 3-30.

BALCAN, D., GONÇALVES, B., HU, H., RAMASCO, J. J., COLIZZA, V. y VESPIGNANI, A. (2010). Modeling the spatial spread of infectious diseases: The Global Epidemic and Mobility Computational Model. *Journal of Computational Science*, 1(3), pp. 132-145.

BRADLEY, V. C., KURIWAKI, S., ISAKOV, M., SEJIDINOVIC, D., MENG, X.-L. y FLAXMAN, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*, 600(7890), pp. 695-700.

CASADO, J. M. y DEL PINO, E. (2021). Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021). *Cuadernos Económicos de ICE*, 102, pp. 13-38.

CASTILLO, J. C., AHUJA, A., ATHEY, S., BAKER, A., BUDISH, E., CHIPTY, T., GLENNERSTER, R., KOMINERS, S. D., KREMER, M., LARSON, G., LEE, J., PRENDERGAST, C., SNYDER, C. M., TABARROK, A., TAN, B. J. y WIĘCEK, W. (2021). Market design to accelerate COVID-19 vaccine supply. *Science*, 371(6534), pp. 1107-1109.

CATALÀ, M., PINO, D., MARCHENA, M., PALACIOS, P., URDIALES, T., CARDONA, P.-J., ALONSO, S., LÓPEZ-CODINA, D., PRATS, C. y ÁLVAREZ-LACALLE, E. (2021). Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers. *PLOS ONE*, 16(1), e0243701.

CHANDRA, A. y SKINNER, J. (2012). Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, 50(3), pp. 645-680.

CHINAZZI, M., DAVIS, J. T., AJELLI, M., GIOANNINI, C., LITVINOVA, M., MERLER, S., PIONTTI, A. P., MU, K., ROSSI, L., SUN, K., VIBOUD, C., XIONG, X., YU, H., HALLORAN, M. E., LONGINI JR, I. M. y VESPIGNANI, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), pp. 395-400.

COMENDEIRO-MAALØE, M., RIDAO-LÓPEZ, M., GORGEMANS, S. y BERNAL-DELGADO, E. (2019). A comparative performance analysis of a renowned public private partnership for health care provision in Spain between 2003 and 2015. *Health Policy*, 123(4), pp. 412-418.

ECDC (2021, septiembre 2022). *Protocol for a focused after-action review on evidence-based decision-making for selected COVID-19 response measures*. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/publications-data/protocol-focused-after-action-review-evidence-based-decision-making-covid-19-response>

EPSTEIN, D. y ESPÍN, J. (2020). Evaluation of new medicines in Spain and comparison with other European

countries. <i>Gaceta Sanitaria</i>, 34, pp. 133-140. EUROPEAN COMMISSION, DG SANTE (2021). <i>The State of Health in the EU: Companion Report 2021</i>. FICHERA, E., MORA, T., LÓPEZ-VALCÁRCEL, B. G. y ROCHE, D. (2021). How do consumers respond to «sin taxes»? New evidence from a tax on sugary drinks. <i>Social Science & Medicine</i>, 274, 113799. FRIEDMAN, J., LIU, P., TROEGER, C. E., CARTER, A., REINER, R. C., BARBER, R. M., COLLINS, J., LIM, S. S., PIGOTT, D. M., VOS, T., HAY, S. I., MURRAY, C. J. L. y GAKIDOU, E. (2021). Predictive performance of international COVID-19 mortality forecasting models. <i>Nature Communications</i>, 12(1), p. 2609. GARCÍA-BASTEIRO, A. L., LEGIDO-QUIGLEY, H., ÁLVAREZ-DARDET, C., ARENAS, A., BENGUA, R., BORRELL, C., VAL, M. D., FRANCO, M., GEA-SÁNCHEZ, M., GESTAL, J., GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., HERNÁNDEZ-AGUADO, I., MARCH, J. C., MARTÍN-MORENO, J., MENÉNDEZ, C., MINUÉ, S., MUNTANER, C., PORTA, M., PRIETO-ALHAMBRA, D. y LEGIDO-QUIGLEY, H. (2020). Evaluation of the COVID-19 response in Spain: Principles and requirements. <i>The Lancet Public Health</i>, 5(11), e575. GARCÍA-SEMPERE, A., ORRICO-SÁNCHEZ, A., MUÑOZ-QUILES, C., HURTADO, I., PEIRÓ, S., SANFÉLIX-GIMENO, G. y Díez-DOMINGO, J. (2020). Data Resource Profile: The Valencia Health System Integrated Database (VID). <i>International Journal of Epidemiology</i>, 49(3), 740-741e. GERTLER, P. J., MARTÍNEZ, S., PREMAM, P., RAWLINGS, L. B. y VERMEERSCH, C. M. J. (2016). <i>Impact Evaluation in Practice</i>. Second Edition. World Bank Publications. GONZÁLEZ LÓPEZ-VALCÁRCEL, B. (2019). Aplicaciones prácticas del método SROI. En: <i>El método SROI en la evaluación económica de intervenciones sanitarias</i>, pp. 69-81. Madrid. Weber. http://weber.org.es/wp-content/uploads/2019/06/Libro-SROI-digital-1.pdf GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., LIBRERO, J., GARCÍA-SEMPERE, A., PEÑA, L. M., BAUER, S., PUIG-JUNOY, J., OLIVA, J., PEIRÓ, S., y SANFÉLIX-GIMENO, G. (2017).	Effect of cost sharing on adherence to evidence-based medications in patients with acute coronary syndrome. <i>Heart</i>, 103(14), pp. 1082-1088. GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., TORRES, L. V. y DOMÍNGUEZ, J. P. (2021). La transformación digital del sector sanitario. <i>Economistas</i>, 174, pp. 75-84. HALE, T., ANGRIST, N., GOLDSZMIDT, R., KIRA, B., PETHERICK, A., PHILLIPS, T., WEBSTER, S., CAMERON-BLAKE, E., HALLAS, L., MAJUMDAR, S. y TATLOW, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). <i>Nature Human Behaviour</i>, 5(4), pp. 529-538. HECKMAN, J. J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. <i>Journal of Political Economy</i>, 109(4), pp. 673-748. HOUSE OF COMMONS (2021). <i>Coronavirus: Lessons learned to date – Health and Social Care, and Science and Technology Committees – House of Commons</i>. https://publications.parliament.uk/pa/cm5802/cmselect/cmsctech/92/9203.htm LINKA, K., PEIRLINCK, M., SAHLI COSTABAL, F. y KUHL, E. (2020). Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. <i>Computer Methods in Biomechanics and Biomedical Engineering</i>, 23(11), pp. 710-717. LÓPEZ, F., CATALÀ, M., PRATS, C., ESTRADA, O., OLIVA, I., PRAT, N., ISNARD, M., VALLÈS, R., VILAR, M., CLOTET, B., ARGIMON, J. M., ARAN, A. y ARA, J. (2022). A Cost&Benefit Analysis of COVID-19 Vaccination in Catalonia. <i>Vaccines</i>, 10(1), p. 59. LÓPEZ-CASASNOVAS, G. y DEL LLANO, J. (2017). <i>Colaboración público-privada en sanidad: El modelo Alzira</i>. Fundacion Gaspar Casal. https://www.riberasalud.com/wp-content/uploads/2019/03/Colaboracion-Publico-Privada-en-Sanidad-El-Modelo-Alzira.pdf MARTÍN-MORENO, J. M., ARENAS, A., BENGUA, R., BORRELL, C., FRANCO, M., GARCÍA-BASTEIRO, A. L., GESTAL, J., GONZÁLEZ	LÓPEZ-VALCÁRCEL, B., HERNÁNDEZ AGUADO, I., LEGIDO-QUIGLEY, H., MARCH, J. C., MINUÉ, S., MUNTANER, C. y VIVES-CASES, C. (2020). Reflexiones sobre cómo evaluar y mejorar la respuesta a la pandemia de COVID-19. <i>Gaceta Sanitaria</i>, 36(1), pp. 32-36. McKEE, M. y STUCKLER, D. (2015). Reflective Practice: How the World Bank Explored Its Own Biases? <i>International Journal of Health Policy and Management</i>, 5(2), pp. 79-82. MILKMAN, K. L., GROMET, D., HO, H., KAY, J. S., LEE, T. W., PANDILOSKI, P., PARK, Y., RAI, A., BAZERMAN, M., BESHEARS, J., BONACORSI, L., CAMERER, C., CHANG, E., CHAPMAN, G., CIALDINI, R., DAI, H., ESKREIS-WINKLER, L., FISHBACH, A., GROSS, J. J. y DUCKWORTH, A. L. (2021). Megastudies improve the impact of applied behavioural science. <i>Nature</i>, 600(7889), pp. 478-483. OCDE (2020a). <i>The Economics of patient safety. From Analysis to Action</i>. https://www.oecd.org/health/health-systems/Economics-of-Patient-Safety-October-2020.pdf OCDE (2020b). <i>Open, Useful and Re-usable data (OURdata) Index: 2019</i>. https://ictlogy.net/bibliography/reports/projects.php?idp=4303 OLIU-BARTON, M. y PRADELSKI, B. S. R. (2021). Green zoning: An effective policy tool to tackle the COVID-19 pandemic. <i>Health Policy</i>, 125(8), pp. 981-986. OREA, L. y ÁLVAREZ, I. C. (2022). How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces. <i>Health Economics</i>, 31(1), pp. 154-173. ORTEGA-DÍAZ, M. I., OCAÑA-RIOLA, R., PÉREZ-ROMERO, C. y MARTÍN-MARTÍN, J. J. (2020). Multilevel Analysis of the Relationship between Ownership Structure and Technical Efficiency Frontier in the Spanish National Health System Hospitals. <i>International Journal of Environmental Research and Public Health</i>, 17(16), p. 5905. POLETTI, C., SCARPINO, S. V. y VOLZ, E. M. (2020). Applications of predictive modelling early in the COVID-19
---	---	---

epidemic. <i>The Lancet Digital Health</i>, 2(10), pp. e498-e499. PÓO, J. M. R. y GALIANO, A. S. (2021). Estadística oficial y evaluación de políticas públicas: Oportunidades y retos. <i>Cuadernos Económicos de ICE</i>, 102, pp. 51-67. RODRÍGUEZ-MIRELES, S., GONZÁLEZ LÓPEZ-VALCÁRCEL, B. y SERRA-MAJEM, L. (2021). When Industrial Policies Conflict With Population Health: Potential Impact of Removing Food Subsidies on Obesity Rates. <i>Value in Health</i>, 24(3), pp. 336-343. ROSSMAN, H. y SEGAL, E. (2022). Nowcasting the spread of SARS-CoV-2. <i>Nature Microbiology</i>, 7(1), pp. 16-17. ROYER, H. (2021). Benefits of megastudies for testing behavioural interventions. <i>Nature</i>, 600(7889), pp. 392-393. SIRLEAF, E. J. y CLARK, H. (2021). Report of the Independent Panel for Pandemic Preparedness and Response: Making	COVID-19 the last pandemic. <i>The Lancet</i>, 398(10295), pp. 101-103. THYGESEN, L. C., CHRISTIANSEN, T., GARCÍA-ARMESTO, S., ANGULO-PUEYO, E., MARTÍNEZ-LIZAGA, N., BERNAL-DELGADO, E. y ECHO CONSORTIUM. (2015). Potentially avoidable hospitalizations in five European countries in 2009 and time trends from 2002 to 2009 based on administrative data. <i>European Journal of Public Health</i>, 25(suppl1), pp. 35-43. UBALDI, B. (2013). <i>Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives</i>. OECD. VALL CASTELLÓ, J. y LÓPEZ CASASNOVAS, G. (2020). Impact of SSB taxes on sales. <i>Economics & Human Biology</i>, 36, 100821. VALLEJO TORRES, L. y GONZÁLEZ LÓPEZ-VALCÁRCEL, B. (2021). The costs of COVID-19 and the cost-effectiveness of testing. <i>Applied Economic Analysis</i>, 29(85), pp. 77-89.	WONG, M. C., HUANG, J., TEOH, J. y WONG, S. H. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: An analysis of 139 countries. <i>The Journal of Infection</i>, 81(3), e70-e71. WORLD HEALTH ORGANIZATION (2019). <i>Guidance for after action review (AAR)</i> (WHO/WHE/CPI/2019.4). World Health Organization. https://apps.who.int/iris/handle/10665/311537 WU, J. T., LEUNG, K., LAM, T. T. Y., NI, M. Y., WONG, C. K. H., PEIRIS, J. S. M. y LEUNG, G. M. (2021). Nowcasting epidemics of novel pathogens: Lessons from COVID-19. <i>Nature Medicine</i>, 27(3), pp. 388-395. ZALA, D., MOSWEU, I., CRITCHLOW, S., ROMEO, R. y McCRONE, P. (2020). Costing the COVID-19 Pandemic: An Exploratory Economic Evaluation of Hypothetical Suppression Policy in the United Kingdom. <i>Value in Health</i>, 23(11), pp. 1432-1437.
---	--	---

LA EVALUACIÓN EN EL CONTEXTO EDUCATIVO

José Manuel CORDERO FERRERA

Francisco PEDRAJA CHAPARRO

Rosa SIMANCAS RODRÍGUEZ

Universidad de Extremadura

Resumen

El presente estudio ofrece un breve repaso de la literatura nacional e internacional centrada en la evaluación de factores vinculados con el rendimiento escolar, prestando especial atención a la descripción de los métodos empleados para realizar dichas evaluaciones y los requisitos de información que requiere su correcta implementación. Asimismo, se presenta un ejemplo práctico de evaluación educativa realizada en la región de Extremadura con el propósito de mejorar la formación de su población a través de un incentivo monetario.

Palabras clave: evaluación, educación, inferencia causal, políticas educativas.

Abstract

This study offers a brief literature review of national and international studies focused on the evaluation of factors related to student achievement, paying special attention to the description of the methods used to conduct such evaluations and the information requirements for their correct implementation. We also present an illustrative example of an educational evaluation conducted in the region of Extremadura with the aim of improving the education level of the population through a monetary incentive.

Keywords: evaluation, education, causal inference, education policies.

JEL classification: I21, H52, C18.

I. INTRODUCCIÓN

JUSTIFICAR la evaluación en educación puede resultar algo obvio. El gasto en educación constituye uno de los programas fundamentales del Estado de bienestar y tiene efectos positivos en los dos objetivos básicos de la economía pública, la eficiencia (crecimiento) y la equidad (igualdad de oportunidades), además de otras consecuencias favorables para la sociedad como son la mejora en los niveles de salud, la reducción de la actividad delictiva o la mejora de su calidad democrática. Por su parte, la evaluación de las políticas públicas constituye un elemento de legitimación de la actuación del sector público en la economía y forma parte de la rendición de cuentas de los Gobiernos a sus ciudadanos. Desde una perspectiva más pragmática, la evaluación constituye un elemento fundamental en la mejora de la gestión pública al basar los programas públicos en función de sus resultados más que en buenas intenciones o meras intuiciones.

Dada la amplitud del contenido, centramos nuestro artículo en la descripción de trabajos de evaluación relacionados con factores explicativos del rendimiento escolar. El interés de la selección se justifica por la relación de estos resultados con el crecimiento económico. Según Hanushek y Woessman (2011), medidas cualitativas de capital humano (resultados escolares en pruebas inter-

nacionales estandarizadas) explican una parte importante de la variación interanual en las tasas de crecimiento del PIB entre países (1).

Los economistas analizamos esta cuestión mediante una función de producción educativa, cuyo análisis empírico se ha visto impulsado y enriquecido tanto por la disponibilidad de datos derivados de la generalización de evaluaciones internacionales de los sistemas educativos como por el desarrollo de métodos econométricos que permiten analizar la amplia variedad de factores que pueden influir en los resultados educativos, que van desde el entorno personal y familiar de los estudiantes y el aula, hasta los recursos escolares y factores de tipo institucional (Todd y Wolpin, 2003; Creemers y Kyriakides, 2007). Dado el espacio del que disponemos, hacemos un repaso ejemplificativo y no exhaustivo de estudios que utilizan técnicas de evaluación de impacto en el análisis de la función de producción educativa desde una doble perspectiva, internacional y española. Continuamos el trabajo con un ejemplo de evaluación educativa, el Programa 18-25, puesto en marcha en Extremadura con la intención de mejorar su capital humano con la obtención de la Enseñanza Secundaria Obligatoria (ESO) ofreciendo un incentivo monetario. Por último, cerramos el artículo con el habitual apartado de conclusiones.

II. LA EVALUACIÓN EN EDUCACIÓN EN PERSPECTIVA INTERNACIONAL

El interés de los economistas por la función de producción educativa tiene su origen en las polémicas conclusiones obtenidas en el famoso informe Coleman (Coleman *et al.*, 1966) cuyo objetivo era comprobar la importancia de los recursos escolares en el éxito escolar, principalmente, en los grupos socioeconómicos más desfavorecidos. Esta investigación, basada en datos de más de 700.000 estudiantes en Estados Unidos, llegó a la conclusión de que los recursos escolares explicaban una parte muy reducida de las diferencias registradas en los resultados escolares en comparación con la importancia que tenían los factores socioeconómicos. Posteriormente, Hanushek (1986), en otro de los trabajos fundamentales dedicados a la función de producción educativa (2), llegó a una conclusión similar, destacando la escasa importancia de ese tipo de recursos.

En buena medida, esos resultados tenían su origen tanto en los datos disponibles como en las técnicas econométricas utilizadas que impedían tener en cuenta la diversidad de factores que forman parte del proceso educativo y la complejidad de sus relaciones. Diversidad y complejidad que dificultaban la identificación de relaciones causales entre las variables analizadas y con ello la posibilidad de evaluar el impacto de las intervenciones en el ámbito educativo.

Para poder evaluar correctamente esas intervenciones es necesario corregir los posibles sesgos que provoca la existencia de endogeneidad en los datos, derivada de la omisión o imposibilidad de observación de factores relevantes, la existencia de causalidad inversa, el sesgo de autoselección o los errores de medición en las variables de interés observadas. La búsqueda de soluciones para evitar los problemas vinculados con la endogeneidad se ha convertido en una de las principales preocupaciones de los investigadores que trabajan la economía de la educación (3). Nuestro propósito en este trabajo no es analizar las técnicas existentes para corregir esos problemas, sino mostrar aquellas que se han utilizado más habitualmente en la literatura de evaluación de políticas educativas, destacando algunas de sus ventajas e inconvenientes, así como los requisitos que exige su correcta implementación.

1. La estrategia de la aleatorización

La mejor manera de corregir los posibles sesgos de selección e identificar los efectos causales es mediante la realización de un experimento aleatorio, es decir, seleccionando aleatoriamente una muestra de la población y, dentro de ella, asignando al azar a los individuos a los grupos de tratamiento y control. De este modo se crean dos grupos con características estadísticamente equivalentes salvo en lo referido a su participación en el programa objeto de evaluación. Esto permitirá concluir que las diferencias encontradas entre los resultados medios del grupo de individuos tratado y del grupo de control se deben al tratamiento (programa), ya que otras posibles explicaciones quedan anuladas por la propia estrategia del proceso de asignación aleatoria (Murnane y Willett, 2011).

Uno de los primeros experimentos aleatorios en el ámbito educativo fue el Proyecto Perry, un programa que se puso en marcha en los años sesenta en Estados Unidos con el propósito de analizar la efectividad de los programas de educación preescolar. Este estudio contó con la participación de 123 alumnos en riesgo de fracaso escolar, a los que se dividió aleatoriamente en dos grupos; uno siguió un programa preescolar de alta calidad basado en el enfoque de aprendizaje activo; y otro, que actuó como grupo de control, no tuvo acceso al programa. La principal particularidad de este estudio es que se hizo un seguimiento longitudinal de los alumnos. Han sido muchos los trabajos que han explotado esta base de datos, aunque los análisis más relevantes, aquellos realizados por el premio nobel de economía James Heckman y sus colaboradores, concluyeron que los beneficiarios del programa obtuvieron mejoras significativas en los ámbitos educativo y laboral (acceso al empleo, estabilidad laboral y salarios), así como mejores niveles de salud y menores tasas de criminalidad (Heckman *et al.*, 2010a; 2010b).

Otro famoso experimento en este campo es el Proyecto STAR (Student/Teacher Achievement Ratio), implementado en los años ochenta en el estado de Tennessee (Estados Unidos) cuyo objetivo fue comprobar el efecto del tamaño de la clase en los resultados escolares y en el que participaron alrededor de 11.600 alumnos. Se trata de un experimento controlado que comenzando en la guardería se extendía hasta tercero de primaria y que asignaba aleatoriamente a los alumnos en clases de tamaño reducido (entre 13 y 17 alumnos) y en

clases de tamaño normal (entre 22 y 25) en las que un ayudante asistía al profesor (Word *et al.*, 1990). Los trabajos que analizaron los resultados del experimento demostraron que aquellos alumnos que asistieron a las clases reducidas consiguieron mejores resultados y tuvieron mayor probabilidad de asistir a la universidad, siendo los efectos mayores en el caso de alumnos pertenecientes a entornos más desfavorecidos (Finn y Achilles, 1999; Chetty *et al.*, 2011).

Con el paso del tiempo, este tipo de experimentos se ha convertido en una práctica más frecuente en la evaluación de las políticas educativas, especialmente en los países anglosajones, donde la cultura de la evaluación está mucho más arraigada. No obstante, también es posible encontrar la aplicación de estas metodologías en países en vías de desarrollo. Los trabajos de Kremer (2003) y Duflo y Kremer (2005) resumen algunas evaluaciones de diferentes intervenciones educativas como son la entrega de dinero para promover la asistencia a la escuela, las comidas gratuitas en centros escolares, la gratuidad de uniformes escolares, la utilización de libros de texto o la incorporación de un segundo maestro a las clases.

Lamentablemente, la realización de estos experimentos se ve limitada por inconvenientes de distinta naturaleza, desde los económicos (por su elevado coste) hasta los éticos (selección de los grupos), pasando por la imposibilidad de su aplicación dado el carácter general de las políticas públicas. En todo caso, parece preferible ensayar las políticas públicas sobre las que existe un alto grado de incertidumbre mediante un programa piloto con el que evaluemos su impacto que aplicar esa política de forma generalizada al conjunto de la población.

Por otra parte, muchas políticas educativas no se diseñan con la intención de ser evaluadas. En este último caso, los investigadores deberán trabajar con los datos observados una vez que la política educativa ha sido implementada, tratando de imitar las condiciones de un experimento aleatorio mediante la utilización de técnicas cuasiexperimentales.

2. Aproximaciones cuasiexperimentales

Estas aproximaciones exigen controlar por todas las diferencias (sean o no observables) existentes entre los grupos de tratamiento y de control para poder estimar correctamente el efecto del tratamiento. Entre ellas cabe destacar el método

de diferencias en diferencias (DiD) y sus variantes (efectos fijos y control sintético), la regresión en discontinuidad (RD) y el enfoque de variables instrumentales (VI) (4)(5). La elección entre estos métodos dependerá de las características del programa a evaluar y del tipo y calidad de los datos disponibles.

La utilización de la estrategia de *diferencias en diferencias* requiere disponer de información sobre los individuos que componen los grupos de tratamiento y control antes y después de la implementación de una determinada intervención, lo que exige la utilización de datos longitudinales. Esto explica que gran parte de los estudios empíricos que aplican esta técnica se hayan llevado a cabo en países desarrollados, donde este tipo de datos suelen estar disponibles gracias a la realización periódica de evaluaciones nacionales a gran escala o donde existen registros administrativos de carácter censal (6).

Un ejemplo puede encontrarse en el trabajo de Machin y McNally (2008) que aplica la técnica de DiD para evaluar el impacto de una hora semanal adicional de lectura introducida en las escuelas de enseñanza primaria de Inglaterra a finales de los años noventa, llegando a la conclusión de que tuvo un impacto positivo. También existen varios trabajos que utilizan esta metodología con datos administrativos para evaluar el impacto de programas de incentivos al profesorado aplicados en diferentes estados americanos (Eberts *et al.*, 2002; Imberman y Lovenheim, 2015) que, en su mayoría, no parecen contribuir significativamente a la mejora del rendimiento de los estudiantes (7). Afortunadamente, en las dos últimas décadas, cada vez hay más países en vías de desarrollo que han hecho el esfuerzo de recopilar microdatos sobre los estudiantes y las escuelas de manera periódica, lo que permite la construcción de bases de datos con estructura de panel y facilita la evaluación de diferentes políticas educativas aplicando la técnica de DiD. Los trabajos de Lai *et al.* (2011) y Ding *et al.* (2020) para China, Angrist *et al.* (2006) o Laajaj *et al.* (2022) con datos colombianos o Glewwe *et al.* (2009) y Lucas y Mbiti (2012) para Kenia son buenos ejemplos.

A pesar de que uno de los requisitos del enfoque DiD es poder observar a los individuos antes y después de la implantación de una determinada política, también es posible aplicarlo con datos de sección cruzada (un único período) siempre que se disponga de, al menos, dos medidas diferentes de la variable de interés para cada individuo de

los grupos de tratamiento y control. Jürges *et al.* (2005), por ejemplo, analizan el efecto causal de la implantación de pruebas estandarizadas de conocimientos en Alemania aprovechando el hecho de que en algunos Estados existen este tipo de pruebas solamente para las matemáticas y no para las ciencias. Por tanto, la primera diferencia es la registrada entre las dos materias y la segunda es la diferencia entre estudiantes en Estados con y sin pruebas. Schwerdt y Wuppermann (2011) y Bietenbeck (2014), por su parte, también analizan el efecto de dos tipologías de prácticas docentes (modernas frente a tradicionales), explotando las diferencias existentes entre los resultados obtenidos en las dos competencias evaluadas en el estudio TIMSS para controlar los rasgos no observados de los estudiantes.

La principal crítica que puede hacerse a estos trabajos es que la variación existente entre los resultados obtenidos por los alumnos en las diferentes materias evaluadas en las pruebas internacionales puede ser ficticia. Esto se debe a que las medidas representativas de esos resultados, los denominados valores plausibles, se generan mediante un proceso de imputación que toma como referencia las respuestas de los estudiantes a un número reducido de preguntas y a su contexto personal y familiar (8). Por tanto, es posible que la variación existente entre asignaturas se explique únicamente por el proceso de imputación utilizado en estas pruebas, especialmente aquellos casos en los que los estudiantes solo hayan respondido a preguntas de una única materia (Jerrim *et al.*, 2017).

La aplicación de la *regresión en discontinuidad* también está condicionada en gran medida por la disponibilidad de un volumen de datos suficiente alrededor del punto de corte en la variable que determina quiénes podrán beneficiarse del programa y quiénes formarán parte del grupo de control. Por esa razón, no es extraño que muchos de los trabajos que emplean este enfoque metodológico utilicen datos administrativos. Este tipo de diseño fue utilizado por primera vez para analizar el efecto de un programa de becas al que accedían aquellos estudiantes que obtuvieran una nota superior a un valor obtenido mediante un test específico diseñado para el programa (Thistlewaite y Campbell, 1960).

Desde entonces, se ha utilizado para analizar políticas educativas muy diversas (9). Esta técnica ha sido profusamente utilizada para analizar el efecto

del tamaño de clase, aprovechando la existencia de normas en muchos países que fijan un tamaño máximo de clase en las escuelas y que obligan a su división en grupos más reducidos en caso de superar ese máximo. En un trabajo reciente, Angrist *et al.* (2019) aplicaron el método de RD a una base de datos de escuelas públicas de Israel y no identificaron ningún efecto significativo de la política de reducción del tamaño de la clase (10). En otro relevante trabajo, Jacob y Lefgren (2009) utilizan esta técnica para analizar el efecto de la política de repetición de curso a partir de datos administrativos de las escuelas públicas de Chicago, llegando a la conclusión de que obligar a repetir a estudiantes con bajo rendimiento en la escuela primaria aumenta sustancialmente la probabilidad de que abandonen la escuela secundaria.

El uso del método de *variables instrumentales* tiene el propósito de eliminar, o al menos mitigar, los problemas de sesgo relacionados con la omisión de variables relevantes o la posible causalidad inversa. Esta alternativa no es tan exigente con los datos necesarios para su aplicación, lo que resulta clave es identificar el instrumento apropiado que pueda ser considerado como una alteración exógena del proceso educativo (11).

Un hecho que puede facilitar la identificación del instrumento es la existencia de una norma o un cambio en la regulación del sistema educativo. Por ejemplo, Machin *et al.* (2007) aprovechan un cambio en las reglas de asignación de recursos para la adquisición de tecnologías de información y conocimiento (TIC) en los distintos distritos escolares de Inglaterra como instrumento para evaluar el efecto causal del gasto en TIC sobre los resultados de los alumnos, identificando un efecto positivo en dos de las tres competencias evaluadas en su estudio.

Otra variable frecuentemente utilizada como instrumento es la fecha de nacimiento de los alumnos, en combinación con la de su incorporación a la escuela, para corregir posibles sesgos asociados por la no consideración de las capacidades innatas de los estudiantes en el análisis. Esta estrategia se ha empleado tanto para analizar el efecto de los años de escolarización (Angrist y Krueger, 1991) como las consecuencias de la política de repetición (Green y Winters, 2007) o simplemente para identificar el efecto de la edad (o el grado de madurez) de los alumnos sobre su rendimiento académico (Lee y Fish 2010).

Por último, una estrategia de identificación habitual consiste en utilizar la localización de las escuelas o la distancia de los alumnos a ellas para estimar su efecto sobre el rendimiento. En particular, este enfoque ha sido principalmente empleado en trabajos que analizan si los alumnos que asisten a escuelas privadas presentan resultados significativa y estadísticamente diferentes de aquellos matriculados en centros de titularidad pública (Vandenberghe y Robin, 2004; Pfeffermann y Landsman, 2011).

En la literatura más reciente podemos encontrar un conjunto de trabajos que utilizan la información que proporcionan profesores y directores de los centros educativos como instrumento para corregir la posible endogeneidad existente en los datos de los estudiantes al evaluar el efecto de una determinada política sobre sus resultados académicos. Por ejemplo, al estudiar los efectos del tiempo dedicado a los deberes, Gustafsson (2013) emplea la información proporcionada por los profesores como instrumento del tiempo que declaran los alumnos dedicar a sus tareas. Sin embargo, el uso de estos instrumentos y, por tanto, los resultados que se derivan de ellos resultan bastante más cuestionables por la dificultad de otorgar un carácter exógeno a las opiniones personales.

III. LA EVALUACIÓN EN EDUCACIÓN EN ESPAÑA

Hasta principios de este siglo, apenas existían en nuestro país trabajos empíricos que analizaran los factores explicativos del rendimiento escolar de los estudiantes debido, en gran medida, a la escasez de información adecuada para realizar este tipo de estudios. Los escasos ejemplos que había estaban, en su mayoría, basados en datos agregados referidos a las escuelas, lo que limitaba el alcance de sus resultados. Sin embargo, desde que España comenzó a participar en las principales evaluaciones internacionales (TIMSS, PIRLS, PISA O PIAAC), se produjo un importante avance en esta línea de investigación gracias a la disponibilidad de una amplia variedad de microdatos sobre múltiples factores que pueden influir en el proceso educativo. Asimismo, la realización de dos evaluaciones nacionales a gran escala (Evaluaciones generales de diagnóstico –EGD–), impulsadas por el Instituto Nacional de Evaluación Educativa (INEE) en primaria (año 2009) y secundaria (año 2010), ofreció a los investigadores dos fuentes de información adicionales para el estudio específico de nuestro sistema educativo.

Al igual que en el contexto internacional, los primeros trabajos que exploraron los determinantes del rendimiento académico se centraron en la identificación de relaciones estadísticamente significativas entre variables (12). Sin embargo, siguiendo la tendencia descrita en el ámbito internacional, pronto comenzaron a surgir trabajos en los que se utilizaban modelos de inferencia causal con el propósito de aproximar las condiciones ideales de los experimentos aleatorios a pesar de que, como se ha mencionado anteriormente, estas bases de datos no están diseñadas para la aplicación de esas técnicas.

Uno de los estudios pioneros en este sentido fue el de Salinas y Santín (2012), en el que se empleaba el uso de variables instrumentales para corregir el posible sesgo en la estimación de la función de producción educativa al comparar el rendimiento de los alumnos inmigrantes y nacionales que asistían a escuelas públicas y concertadas. Una de sus principales conclusiones fue que la concentración de alumnos inmigrantes en centros públicos repercutía negativamente sobre los nativos, mientras que una baja concentración de aquel tipo de alumnado en las escuelas concertadas mejoraba el rendimiento educativo de los mismos. A una conclusión similar llegaron Pedraja *et al.* (2016) al analizar el efecto de la concentración de inmigrantes en determinados centros educativos utilizando datos procedentes de la base de datos PISA, aunque en su caso la estrategia de estimación se basaba en el uso de diferencias en diferencias.

El efecto de la política de repetición de curso sobre el rendimiento escolar de los alumnos es otra de las cuestiones que ha atraído la atención de los investigadores españoles debido a que nuestro país alcanza una de las mayores tasas de repetición de Europa (López-Rupérez *et al.*, 2021). La obtención de resultados fiables exige tener en cuenta el problema de causalidad inversa, ya que si los repetidores presentan peores resultados puede ser debido a la repetición o a que ya tenían un peor rendimiento antes de repetir. Los únicos estudios empíricos que han empleado enfoques metodológicos que permiten corregir posibles sesgos derivados de este problema han sido los de García-Pérez *et al.* (2014), mediante el uso de variables instrumentales, y Choi *et al.* (2018), incorporando el rendimiento previo de los alumnos mediante la construcción de un pseudopanel. Ambos llegan a la conclusión de que la estrategia de repetición tiene un efecto negativo sobre el rendimiento, siendo este mucho más grave en primaria que en secundaria.

También podemos encontrar estudios que han tratado de establecer una relación causal entre la asistencia a la educación preescolar y los resultados educativos, intentando corregir posibles sesgos provocados por la endogeneidad implícita de esta decisión (los años de asistencia a educación infantil están correlacionados con variables no observables como sucede con la motivación de los padres), mediante diferentes aproximaciones cuasiexperimentales. En todos ellos se llega a la conclusión de que existe una relación estadísticamente significativa y positiva entre ambas variables (Santín y Sicilia, 2015; Mancebón *et al.*, 2018). Asimismo, Felfe *et al.* (2015) analizaron el impacto que tuvo la ampliación gratuita de la educación infantil a los tres años con la implantación de la LOGSE (Ley de Ordenación General del Sistema Educativo) a principios de los años noventa, explotando el experimento natural que se produjo como consecuencia de la distinta extensión de la reforma en las comunidades autónomas. Utilizando datos de diferentes oleadas de PISA y un enfoque de diferencias en diferencias, se comprobó que la reforma afectó positivamente a la promoción de curso durante la educación primaria y supuso una mejora en los resultados educativos en lectura y matemáticas que se mantuvo en el largo plazo. Además, este efecto fue mucho más acusado en los alumnos con peor nivel socioeconómico.

También contamos con estudios que han explotado la información disponible en las bases de datos internacionales y nacionales para analizar el efecto de diferentes factores como el número de horas de clase impartidas o las estrategias docentes que emplean los profesores. Sobre el primer aspecto, López-Agudo y Marcenaro (2019a, 2019b), tras aplicar un modelo de efectos fijos a los datos disponibles en PISA, obtienen que el tiempo de instrucción semanal no parece afectar al rendimiento académico. En cuanto al segundo, la efectividad de métodos de enseñanza más modernos frente a los tradicionales, la evidencia no es concluyente ya que mientras Hidalgo y López-Mayán (2018), con datos procedentes de la EGD y aplicando un enfoque DiD, encuentran un efecto positivo sobre el rendimiento académico, Cordero y Gil-Izquierdo (2018), con datos procedentes de PISA y aplicando variables instrumentales, obtienen un efecto negativo. Sin embargo, ambos estudios coinciden en la identificación de un efecto positivo en el uso de las estrategias docentes tradicionales.

Un aspecto que ha suscitado gran interés acrecentado por la actual pandemia, ha sido el impacto

de la incorporación de las TIC en las escuelas a través de diferentes programas impulsados tanto por el Ministerio de Educación como por las comunidades autónomas. Lamentablemente, la mayoría de estas iniciativas no obedecieron a un plan consensuado para todo el territorio nacional ni establecieron un criterio de evaluación riguroso, por lo que resulta muy difícil comprobar si resultaron ser eficaces. La excepción ha sido el programa Escuela 2.0, cuyo objetivo era mejorar la dotación de ordenadores de los centros educativos y para el que existen estudios que han analizado su impacto mediante la aplicación de técnicas de inferencia causal a los datos disponibles en PISA (Villaplana, 2014; Cabras y Tena, 2016). Ambos concluyen que el efecto fue moderadamente positivo solamente en el caso de que tal política viniera acompañada con cambios en la metodología docente del profesorado.

Al margen de los trabajos mencionados, todos ellos basados en información procedente de bases de datos internacionales, existen otros que han explotado la información de la que disponen las comunidades autónomas gracias a las evaluaciones que estas realizan periódicamente en su territorio. Su carácter censal y estructura longitudinal se adapta mucho mejor a la utilización de técnicas de inferencia causal. Así, podemos encontrar estudios que explotan los datos administrativos de los estudiantes de Andalucía para analizar cómo afecta al rendimiento escolar el tamaño de clase (López-Agudo y Marcenaro, 2021) o el tiempo que dedican a la realización de deberes y a la lectura (Jerrim *et al.*, 2019; 2020). También se pueden encontrar trabajos que analizan los datos censales de los estudiantes madrileños para examinar la efectividad de los programas de bilingüismo implantados en esa Comunidad (Anghel *et al.*, 2016; Mediavilla *et al.*, 2019) o los datos relativos a estudiantes catalanes para evaluar el impacto del Programa para la Mejora de la Calidad Educativa implantado en varios centros educativos de Cataluña (López-Torres *et al.*, 2019). Por último, los microdatos de los estudiantes de primaria y secundaria del País Vasco han sido utilizados como referencia para analizar la pérdida de aprendizaje provocada por el cierre de los colegios por la pandemia de la COVID-19 (Arenas y Gortazar, 2022).

Lamentablemente, estas bases de datos son de carácter informativo e interno y no están disponibles al público en general, de manera que su explotación se encuentra supeditada a la voluntad de que la institución responsable de los datos los

facilite a los investigadores, lo que solo sucede de forma excepcional. Esta situación debería corregirse pues son muchos los recursos destinados a recopilar esa información que debería estar disponible de modo general. Como dijimos, la evaluación de las políticas constituye un elemento de legitimación de la actuación del sector público en la economía y forma parte de la rendición de cuentas de los Gobiernos a sus ciudadanos.

La gran asignatura pendiente en nuestro país es el desarrollo de experimentos aleatorios para llevar a cabo la evaluación de las políticas educativas. Es cierto que existen algunas excepciones dignas de mención, como el estudio piloto desarrollado por el Banco de España para evaluar la efectividad de la implantación de un programa de educación financiera (Bover *et al.*, 2018) o el proyecto financiado por el programa Erasmus+ de la Unión Europea para medir hasta qué punto la educación cívica activa influye en la participación y compromiso democrático de los jóvenes mediante una intervención experimental a gran escala en la que participaron Francia, Inglaterra, Grecia y España (Briole *et al.*, 2022). Por otro lado, cabe destacar la labor llevada a cabo en los últimos años por la Fundación «la Caixa», que se ha decidido a promocionar la innovación educativa basada en evidencias poniendo en marcha el programa «Evaluación de programas educativos» de EduCaixa, cuyo objetivo es fomentar la evaluación de intervenciones e innovaciones educativas poniendo en contacto centros educativos con propuestas educativas innovadoras y equipos de investigación (13).

IV. UN EJEMPLO DE EVALUACIÓN EDUCATIVA EN ESPAÑA: EL PROGRAMA 18-25

En este apartado se ofrece un breve resumen de un estudio empírico realizado en el contexto español con el propósito de mejorar el nivel de formación de la población desempleada. Se trata de una iniciativa puesta en marcha en Extremadura, donde en el año 2012 el 41 por 100 de los desempleados inscritos en el Servicio Extremeño Público de Empleo carecían de la Educación Secundaria Obligatoria (ESO). Con el propósito de reducir esas cifras, el Gobierno autonómico aprobó, en noviembre de ese mismo año, el denominado Programa 18-25, el cual se mantuvo vigente durante tres cursos académicos, suprimiéndose con el cambio de Gobierno regional en mayo de 2015.

El objetivo de este programa era reducir, mediante un incentivo económico de 1.000 euros, el número de desempleados extremeños que carecían de la formación básica obligatoria y con ello mejorar sus posibilidades de incorporación al mundo laboral. Los beneficiarios potenciales eran todos los desempleados con edades comprendidas entre 18 y 25 años que no contasen con el título de la ESO. Además, en el caso de las mujeres, también podían beneficiarse las mayores de 25 años que se encontrasen en situación de desempleo de larga duración. El programa se gestionaba de acuerdo con lo establecido en la Educación Secundaria para Adultos (ESPA) en Extremadura y, por tanto, estaba organizado en módulos en lugar de asignaturas. El alumno interesado en acogerse al Programa 18-25 debía indicarlo expresamente en su solicitud de admisión en formación de adultos, matriculándose del total de módulos pendientes hasta un máximo de seis (que era el número máximo de módulos permitidos por curso académico en la ESPA). El incentivo económico estaba asociado al rendimiento académico, de manera que los participantes en el programa debían asistir regularmente a clase y aprobar los módulos en los que estuvieran matriculados. Así, el alumno que cumpliera los requisitos anteriores recibiría un pago de 500 euros al finalizar cada cuatrimestre.

Con esta evaluación se pretende comprobar si el programa incrementó la probabilidad de obtener el título de formación básica. Para ello, compararemos el grado de éxito de aquellos que se acogieron al mismo con los que, asistiendo durante el mismo curso a la ESPA, no pudieron beneficiarse del programa por no cumplir el requisito de la edad (ser mayores de 25 años). Este análisis se centrará en la segunda convocatoria del programa, la correspondiente al curso 2013-2014 y, puesto que no es posible evaluar el efecto del programa sobre toda la población objetivo por tener requisitos diferentes según género, optamos por observar su impacto únicamente sobre la población masculina.

Dentro de las diferentes técnicas cuasiexperimentales, se seleccionó como herramienta de análisis la regresión en discontinuidad por ser la que mejor se ajustaba a la naturaleza del programa que, al impedir el acceso a los mayores de 25 años, provoca una discontinuidad en el tratamiento. Este enfoque permite comparar a individuos en un entorno cercano al punto de corte con características razonablemente similares, lo que facilita la identificación del efecto causal de la política.

Las regresiones en discontinuidad se aplican en aquellas situaciones en las que los beneficiarios y no beneficiarios de una política específica son determinados por el hecho de encontrarse por encima o por debajo de un valor concreto $\bar{x}$ de una variable de asignación X_i . La principal ventaja de las regresiones en discontinuidad es que, al comparar los resultados obtenidos por unidades situadas en un entorno cercano (por encima y por debajo) del punto de corte, nos acerca a lo conocido como asignación aleatoria al tratamiento y, por tanto, cualquier diferencia en los resultados puede interpretarse como efecto causal del programa (Gertler *et al.*, 2016). La regresión en discontinuidad puede ser estricta (cuando la variable de asignación determina los grupos de tratamiento y control de forma inequívoca) o difusa (la variable de asignación no permite determinar con exactitud la participación en el programa, sino más bien la probabilidad de participar en el mismo [Schlotter *et al.*, 2011]). Esto último sucede en nuestro caso, donde existen individuos que, cumpliendo los requisitos, decidieron no participar en el programa mientras que otros individuos, aun no cumpliendo el requisito de la edad, acabaron formando parte del grupo de tratados. Esto significa que la variable de asignación (Edad) no determinó exactamente la participación, lo que nos llevó a estimar una regresión discontinua difusa.

Las regresiones discontinuas difusas pueden interpretarse desde un enfoque de variables instrumentales, donde, a partir de la variable de asignación X_i , se genera una variable I_i , que funciona como instrumento de la variable tratamiento D_i (Angrist y Pischke, 2014). De este modo, las ecuaciones estimadas son las siguientes:

1.^a etapa o ecuación de tratamiento:

$$D_i = \gamma_0 + \gamma_1 I_i + \gamma_2 X_i + \gamma_3 Z_i + \varepsilon_1 \quad (1)$$

2.^a etapa o ecuación de resultado:

$$Y_i = \beta_0 + \beta_1 \hat{D}_i + \beta_2 X_i + \beta_3 Z_i + \varepsilon_2 \quad (2)$$

donde Y_i representa la medida del impacto del programa (en nuestro caso, toma valor 1 si, al finalizar el curso académico, el individuo consiguió el título de la ESO y 0 en caso contrario); D_i indica el tratamiento real (variable dicotómica que toma valor 1 si el alumno participó en el Programa 18-25 y 0 en caso contrario); $\hat{D}_i$ representa el tratamiento estimado; I_i es el instrumento empleado para dicha estimación (es igual a 1 si el individuo tenía una

edad igual o inferior a 25 años y 0 en caso contrario); X_i es la variable de asignación (edad del alumno a 31 de diciembre de 2013); y Z_i recoge las variables de control (como características socioeconómicas se incorporan los ingresos medios de la unidad familiar; el entorno del alumno se delimita mediante la tasa de desempleo del municipio donde el alumno asiste a clase y si vive en una localidad considerada rural; por último, se agrega información sobre el número de módulos en los que está matriculado).

Para llevar a cabo la evaluación hemos contado con información obtenida a partir de encuestas realizadas durante las primeras semanas del curso a los alumnos matriculados en educación de adultos. Simultáneamente, el Servicio de Enseñanza de Personas Adultas y a Distancia del Gobierno de Extremadura nos proporcionó información administrativa complementaria. Una vez fusionada la información disponible, restringimos la muestra para centrar el análisis en: i) la población masculina; ii) matriculada en al menos cuatro módulos; y iii) con edades comprendidas entre 20 y 31 años.

El cuadro n.º 1 recoge los resultados de los dos modelos estimados para estudiar el impacto del Programa 18-25. El Modelo 1 no tiene en cuenta las variables de control, mientras que el Modelo 2 sí las incluye para reducir los posibles sesgos derivados de la amplitud de la horquilla. Además, puesto que dicha amplitud (seis años a cada lado del punto de corte) puede dar lugar a que la muestra esté formada por personas con características muy diferentes (individuales, sociales y económicas), repetimos el análisis reduciendo la horquilla, en primer lugar, a cuatro años por encima y por debajo del punto de corte (todos aquellos con edades comprendidas entre los 22 y 29 años) y, en segundo lugar, a los varones entre 24 y 27 años (dos años a cada lado del punto de corte).

De acuerdo con el Modelo 1, el tratamiento resulta no ser estadísticamente significativo en ninguna de las horquillas de edad, observándose resultados similares en el Modelo 2, por lo que podemos concluir que, incluso corrigiendo por el posible sesgo existente mediante la incorporación de variables de control, el impacto del tratamiento continúa siendo no estadísticamente significativo. En definitiva, los resultados muestran que la probabilidad de obtener el título de la ESO no fue estadísticamente diferente para el grupo de tratados y para el grupo de control y, en consecuencia, el hecho de participar en el Programa 18-25 no supu-

CUADRO N.º 1

IMPACTO DEL PROGRAMA 18-25 SOBRE LA PROBABILIDAD DE TITULAR

HORQUILLA	DE 24 A 27 AÑOS		DE 22 A 29 AÑOS		DE 20 A 31 AÑOS	
VARIABLES	MODELO 1	MODELO 2	MODELO 1	MODELO 2	MODELO 1	MODELO 2
Tratamiento	0,6102 (0,4992)	0,4830 (0,3945)	-0,0306 (0,2332)	-0,0442 (0,2232)	-0,1100 (0,1910)	-0,1285 (0,1913)
Edad	0,2556* (0,1329)	0,2424** (0,1099)	0,0222 (0,0318)	0,0223 (0,0325)	0,0051 (0,0186)	0,0054 (0,0192)
Ingresos 1.000-1.500€		0,1841 (0,1528)		-0,0177 (0,1111)		0,0057 (0,0865)
Ingresos > 1.500€		0,4966*** (0,1864)		0,0981 (0,1812)		-0,0735 (0,1357)
Tasa desempleo		0,0263*** (0,0100)		0,0040 (0,0074)		0,0063 (0,0064)
Rural		-0,1042 (0,1085)		0,0106 (0,0750)		-0,0258 (0,0617)
Módulos matriculados		-0,1768** (0,0802)		-0,1083* (0,0615)		-0,0687 (0,0496)
Constante	-6,4226* (3,6319)	-5,7688* (2,8337)	-0,1761 (0,9049)	0,3143 (0,9476)	0,2851 (0,5493)	0,5109 (0,5938)
Observaciones	100		190		279	

Nota: Errores estándar entre paréntesis. ***significativo al 1 %; **5 %, *10 %.

Fuente: Elaboración propia.

so, para los hombres comprendidos en las edades evaluadas, un incentivo suficiente.

V. CONCLUSIONES

El análisis realizado ha tratado de poner de relieve algunas de las principales aproximaciones que habitualmente se utilizan en el ámbito educativo para evaluar políticas o intervenciones dirigidas a mejorar los resultados académicos de los alumnos, como requisito previo para poder contar en el futuro con una población más y mejor formada que garantice mayores tasas de crecimiento y desarrollo económico.

El método ideal para realizar evaluaciones de políticas educativas es el desarrollo de experimentos aleatorios que permiten comprobar si la implementación de una medida ha tenido los efectos deseados mediante la comparación de resultados entre aquellos a los que se aplica la política educativa (grupo tratado) y a los que no (grupo de control). Su uso es cada vez más habitual, especialmente en algunos países anglosajones en los que su aplicación se ha convertido en requisito a la hora de implementar tales políticas. En los últimos años resulta

más frecuente encontrar ejemplos de su empleo también en países en vías de desarrollo, llegando la labor de tres investigadores en este campo (Esther Duflo, Abhijit Banerjee y Michael Kremer) a ser reconocida en 2019 con un Premio Nobel por su labor de fomento del uso de ensayos aleatorios en aquellos países. En España, sin embargo, la utilización de este tipo de diseños es todavía una asignatura pendiente, aunque hay que reconocer su promoción por parte de algunas instituciones privadas que han puesto en marcha algunas pruebas piloto descritas en este artículo. En todo caso, es preciso una mayor implicación de las instituciones públicas en la necesidad de evaluar las intervenciones con el fin de mejorar su diseño. El papel que en la actualidad está desarrollando la Autoridad Independiente de Responsabilidad Fiscal (AIReF) con la creación este año de la división de evaluación del gasto público va en la buena dirección.

Cuando no es posible utilizar experimentos aleatorios por razones diversas, fundamentalmente éticas y económicas, los investigadores recurren al uso de técnicas cuasiexperimentales con las que tratan de reproducir las condiciones ideales de los experimentos aleatorios corrigiendo

los habituales problemas de endogeneidad. Su correcta implementación se facilita si se dispone de datos administrativos o censales (Connelly *et al.*, 2016) con los que poder analizar a toda la población de interés aprovechando las variaciones existentes entre los diferentes grupos que componen la base de datos, o para crear cohortes de individuos con el fin de estudiar los cambios que se producen entre dos períodos y reunir información sobre individuos que han vivido un determinado cambio (repetición de curso, modificación en la regulación del sistema educativo en diferentes zonas geográficas, etc.).

En España, la responsabilidad de recopilar y codificar este tipo de datos sobre los alumnos de educación primaria y secundaria de manera periódica recae sobre las comunidades autónomas, las responsables de su prestación. Su explotación en estudios empíricos, con técnicas de inferencia causal, es todavía muy escasa ya que no están disponibles, con generalidad, para la comunidad académica por los impedimentos que ponen tales administraciones educativas.

Una alternativa a los datos de carácter censal en la evaluación de políticas educativas son los datos muestrales proporcionados por diferentes evaluaciones nacionales e internacionales de conocimientos en distintas materias a gran escala, una práctica que se ha vuelto cada vez más habitual tanto en los estudios desarrollados en el ámbito internacional como especialmente en el caso de España, dada la ausencia o las dificultades de acceso a mejores fuentes de información. A pesar de que estos datos no fueron diseñados originalmente para la implementación de técnicas de inferencia causal, en la literatura pueden encontrarse múltiples ejemplos en los que los investigadores tratan de estimar efectos causales de programas o intervenciones educativas a partir de ese tipo de datos (Cordero *et al.*, 2018).

Por último, en relación con el ejercicio de evaluación presentado, el Programa 18-25 implementado por el Gobierno de Extremadura a principios de la década anterior con el propósito de mejorar el nivel de formación de los desempleados mediante un incentivo económico, los resultados obtenidos aplicando la técnica de regresión en discontinuidad muestran que la probabilidad de obtener el título de la ESO no fue estadísticamente diferente para el grupo de individuos que participaron en el programa respecto a los que no participaron. Por tanto,

se puede concluir que el Programa 18-25 no resultó efectivo. Este resultado es similar al obtenido en otra investigación realizada por los mismos autores con un diseño similar, aunque con alguna variación en los datos empleados (Pedraja *et al.*, 2022).

NOTAS

(1) Con resultados del informe PISA, aproximadamente tres cuartas partes de la variación de las tasas de crecimiento entre países pueden explicarse por dos variables: el nivel inicial de ingresos y los resultados académicos de la población.

(2) Este artículo resume los principales resultados obtenidos en más de 100 estudios sobre la función de producción educativa realizados en Estados Unidos.

(3) Los interesados en los problemas que origina la endogeneidad en la estimación de modelos econométricos en el ámbito educativo y en los principales métodos experimentales y cuasiexperimentales para la identificación de efectos causales, pueden consultar los trabajos de WEBBINK (2005) o SCHLOTTER *et al.* (2011).

(4) Una detallada explicación de estas técnicas se encuentra en el trabajo de ARTÉS y RODRÍGUEZ-SÁNCHEZ (2022) en este mismo número o en alguno de los manuales sobre evaluación de impacto existentes en la literatura, entre otros, ANGRIST y PISCHKE (2014), KHANDKER *et al.* (2010) o GERTLER *et al.* (2016).

(5) No incluimos dentro de esta categoría a los modelos de *matching* o emparejamiento (entre los que se incluye el *propensity score matching*) porque, aunque reducen los sesgos que genera el problema de la auto-selección, no corrigen los sesgos asociados a la existencia de diferencias no observadas entre el grupo tratado y el de control.

(6) En Estados Unidos existe una Evaluación Nacional del Progreso Educativo (NAEP, por sus siglas en inglés) desde 1969. Además, desde mediados de los años noventa, muchos Estados disponen de sistemas de evaluación periódica de los conocimientos de los estudiantes que permiten un seguimiento longitudinal y emparejarlos con sus escuelas y, en algunos casos, con sus profesores. Los tres Estados que cuentan con los sistemas más antiguos y que han estado ampliamente disponibles para los investigadores son Florida, Carolina del Norte y Texas. En Europa, los países nórdicos cuentan con registros administrativos desde mediados de los años sesenta, en Reino Unido existe una base de datos censal (NPD) desde mediados de los noventa, y en Italia (INVALSI) desde 2005.

(7) Sin embargo, la mayoría de los programas de incentivos al profesorado que se han instaurado en países en vías de desarrollo, donde el margen de mejora es mucho más amplio y los salarios de los docentes son más bajos, han tenido un impacto muy positivo, consiguiendo generar mejoras importantes en los resultados de los estudiantes (GANIMIAN y MURNAME, 2016).

(8) Véase MISLEVY *et al.* (1992) o WU (2005) para una explicación detallada del concepto de valores plausibles y los procedimientos empleados para su construcción.

(9) Los trabajos de VAN DER KLAUW (2008) y CATTANEO *et al.* (2018) ofrecen explicaciones detalladas de este enfoque metodológico y muchos ejemplos prácticos sobre la aplicación de la regresión en discontinuidad.

(10) Este trabajo es una actualización de otro muy influyente en la literatura publicado veinte años antes (ANGRIST y LAVY, 1999), en el que, utilizando una estrategia de identificación muy parecida, los autores sí encontraron un efecto positivo de la reducción del tamaño de clase sobre los resultados escolares.

(11) Los trabajos de MURNANE y WILLET (2011) y POKROPEK (2016) ofrecen una excelente discusión acerca de esta cuestión y clasifican las fuentes más populares de posibles instrumentos de investigación en el ámbito educativo.

(12) Para una revisión de los determinantes del rendimiento académico a partir de los datos de PISA, véase CORDERO *et al.* (2013).

(13) Para más información, véase: <https://educaixa.org/es/evaluacion-programas-educativos>

BIBLIOGRAFÍA

ANGHEL, B., CABRALES, A. y CARRO, J. M. (2016). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry*, 54(2), pp. 1202-1223.

ANGRIST, J., BETTINGER, E. y KREMER, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3), pp. 847-862.

ANGRIST, J. D. y LAVY, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2), pp. 533-575.

ANGRIST, J. D., LAVY, V., LEDER-LUIS, J. y SHANY, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3), pp. 309-324.

ANGRIST, J. D. y KEUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), pp. 979-1014.

ANGRIST, J. D. y PISCHEKE, J. S. (2014). *Mastering metrics: The path from cause to effect*. Princeton University Press.

ARENAS, A. y GORTAZAR, L. (2022). *Learning loss One Year After School Closures: Evidence from the Basque Country*. Working Paper #1, Esade Center for Economic Policy.

ARTÉS, J. y RODRÍGUEZ-SÁNCHEZ, B. (2022). Métodos de evaluación de políticas públicas. *Papeles de Economía Española*, 172, pp. 18-29..

BIETENBECK, J. (2014). Teaching practices and cognitive skills. *Labour Economics*, 30, pp. 143-153.

BOVER, O., HOSPIDO, L. y VILLANUEVA, E. (2018). The impact of high school financial education on financial knowledge and choices: Evidence from a randomized trial in Spain. *IZA Discussion Papers*, n.º 11265. Institute of Labor Economics (IZA), Bonn.

BRIOLE, S., GURGAND, M., MAURIN, É., McNALLY, S., RUIZ-VALENZUELA, J. y SANTÍN, D. (2022). The Making of Civic Virtues: a school-based experiment in three countries. *IZA Discussion Papers*, n.º 15141. Institute of Labor Economics (IZA), Bonn.

CABRAS, S. y TENA, J. D. (2016). A Bayesian non-parametric modeling to estimate student response to ICT investment. *Journal of Applied Statistics*, 43(14), pp. 2627-2642.

CATTANEO, M. D., IDROBO, N. y TITIUNIK, R. (2018). *A Practical Introduction to Regression Discontinuity Designs, Vol. I y II*. Cambridge: Cambridge University Press.

CHETTY, R., FRIEDMAN, J. N., HILGER, N., SAEZ, E., SCHANZENBACH, D. W. y YAGAN, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4), pp. 1593-1660.

CHOI, Á., GIL, M., MEDIAYILLA, M. y VALBUENA, J. (2018). The evolution of educational inequalities in Spain: dynamic evidence from repeated cross-sections. *Social Indicators Research*, 138(3), pp. 853-872.

COLEMAN, J., CAMPBELL, E., HOBSON, C., MC PARTLAND, I., MODD, A., WENFELD, F. y YORK, R. (1966). *Equality of educational opportunity*. Department of Health, Education and Welfare, Office of Education, Government Printing Office, Washington.

CONNELLY, R., PLAYFORD, C. J., GAYLE, V. y DIBBEN, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, pp. 1-12.

CORDERO, J. M., CRESPO-CEBADA, E. y PEDRAJA CHAPARRO, F. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de Educación*, 362, pp. 273-297.

CORDERO, J. M., CRISTÓBAL, V. y SANTÍN, D. (2018). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys*, 32(3), pp. 878-915.

CORDERO, J. M. y GIL-IZQUIERDO, M. (2018). The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *Journal of Policy Modeling*, 40(6), pp. 1313-1331.

CREEMERS, B. y KYRIAKIDES, L. (2007). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

DING, Y., LU, F. y YE, X. (2020). Intergovernmental transfer under heterogeneous accountabilities: the effects of the 2006 Chinese education finance reform. *Economics of Education Review*, 77, 101985.

DUFLO, E. y KREMER, M. (2005). Use of Randomization in the Evaluation of Development Effectiveness. En O. FEINSTEIN, G. K. INGRAM y G. K. PITMAN (eds.), *Evaluating development effectiveness*, pp. 205-232. New Brunswick, New Jersey and London: Transaction Publishers.

EBERTS, R., HOLLENBECK, K. y STONE, J. (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources*, 37(4), pp. 913-927.

FELFE, C., NOLLENBERGER, N. y RODRÍGUEZ-PLANAS, N. (2015). Can't buy mommy's love? universal childcare and children's long-term cognitive development. *Journal of Population Economics*, 28(2), pp. 393-422.

FINN, J. D. y ACHILLES, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), pp. 97-109.

GANIMIAN, A. J. y MURNANE, R. J. (2016). Improving education in developing countries: Lessons from rigorous impact

evaluations. *Review of Educational Research*, 86(3), pp. 719-755.

GARCÍA-PÉREZ, J. I., HIDALGO-HIDALGO, M. y ROBLES-ZURITA, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, 46, pp. 1373-1392.

GERTLER, P. J., MARTÍNEZ, S., PREMAND, P., RAWLINGS, L. B. y VERMEERSCH, C. M. (2016). *Impact evaluation in practice*. The World Bank.

GLEWWE, P., KREMER, M. y MOULIN, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied economics*, 1(1), pp. 112-135.

GREENE, J. P. y WINTERS, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, 2(4), pp. 319-340.

GUSTAFSSON, J. E. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement*, 24(3), pp. 275-295.

HANUSHEK, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), pp. 1141-1177.

HANUSHEK, E. A. y WOESSMANN, L. (2011). The economics of international differences in educational achievement. En E. HANUSHEK, S. MACHIN y L. WOESSMANN (eds.), *Handbook of the Economics of Education*, 3, pp. 89-200. Amsterdam: North Holland.

HECKMAN, J. J., MOON, S. H., PINTO, R., SAVELYEV, P. A. y YAVITZ, A. (2010a). The rate of return to the High Scope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), pp. 114-128.

HECKMAN, J., MOON, S. H., PINTO, R., SAVELYEV, P. y YAVITZ, A. (2010b). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, 1(1), pp. 1-46.

HIDALGO, A. y LOPEZ-MAYAN, C. (2018). Teaching styles and achievement: Student and teacher perspectives. *Economics of Education Review*, 67, pp. 184-206.

IMBERMAN, S. A. y LOVENHEIM, M. F. (2015). Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics*, 97(2), pp. 364-386.

JACOB, B. A. y LEFGREN, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3), pp. 33-58.

JERRIM, J., LÓPEZ-AGUDO, L. A., MARCENARO-GUTIÉRREZ, O. D. y SHURE, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, 61, pp. 51-58.

JERRIM, J., LÓPEZ-AGUDO, L. A. y MARCENARO-GUTIÉRREZ, O. D. (2019). The relationship between homework and the academic progress of children in Spain during compulsory elementary education: A twin fixed-effects approach. *British Educational Research Journal*, 45(5), pp. 1021-1049.

JERRIM, J., LÓPEZ-AGUDO, L. A. y MARCENARO-GUTIÉRREZ, O. D. (2020). Does it matter what children read? New evidence using longitudinal census data from Spain. *Oxford Review of Education*, 46(5), pp. 515-533.

JÜRGES, H., SCHNEIDER, K. y BÜCHEL, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of the European Economic Association*, 3(5), pp. 1134-1155.

KHANDKER, S. R., KOOLWAL, G. B. y SAMAD, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. Washington DC: World Bank.

KREMER, M. (2003). Randomized evaluations of educational programs in developing countries: Some lessons. *American Economic Review*, 93(2), pp. 102-106.

LAAJAJ, R., MOYA, A. y SÁNCHEZ, F. (2022). Equality of opportunity and human capital accumulation: Motivational effect of a nationwide scholarship in Colombia. *Journal of Development Economics*, 154, 102754.

LAI, F., SADOULET, E. y DE JANVRY, A. (2011). The contributions of school quality and teacher qualifications to student performance evidence from a natural experiment in Beijing middle schools. *Journal of Human Resources*, 46(1), pp. 123-153.

LEE, J. y FISH, R. M. (2010). International and interstate gaps in value-added math achievement: Multilevel instrumental variable analysis of age effect and grade effect. *American Journal of Education*, 117(1), pp. 109-137.

LÓPEZ-AGUDO, L. A. y MARCENARO-GUTIÉRREZ, O. D. (2019a). The Effect of Weekly Instruction Time on Academic Achievement: The Spanish Case. *Hacienda Pública Española*, 230, pp. 63-93.

LÓPEZ-AGUDO, L. A. y MARCENARO-GUTIÉRREZ, O. D. (2019b). Are Spanish children taking advantage of their weekly classroom time? *Child Indicators Research*, 12(1), pp. 187-211.

LÓPEZ-AGUDO, L. A. y MARCENARO-GUTIÉRREZ, O. D. (2021). La falta de influencia del tamaño de la clase sobre el rendimiento académico de los estudiantes: evidencia empírica para Andalucía. *Revista de Educación*, 395, pp. 321-361.

LÓPEZ-RUPÉREZ, F., GARCÍA-GARCÍA, I. y EXPÓSITO, E. (2021). La repetición de curso y la graduación en Educación Secundaria Obligatoria en España: análisis empíricos y recomendaciones políticas. *Revista de Educación*, 394, pp. 325-353.

LÓPEZ-TORRES, L., PRIOR, D. y SANTÍN, D. (2019). Assessing the effect of educational programs on public schools' performance. *Applied Economics*, 51(48), pp. 5205-5226.

- LUCAS, A. M. y MBITI, I. M. (2012). Access, sorting, and achievement: The short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics*, 4(4), pp. 226-253.
- MACHIN, S. y McNALLY, S. (2008). The literacy hour. *Journal of Public Economics*, 92(5-6), pp. 1441-1462.
- MACHIN, S., McNALLY, S. y SILVA, O. (2007). New technology in schools: Is there a payoff? *The Economic Journal*, 117(522), pp. 1145-1167.
- MANCEBÓN, M. J., PÉREZ-XIMÉNEZ DE EMBÚN, D. y VILLAR-ALDONZA, A. (2018). Evaluación del efecto de la escolarización temprana sobre las habilidades cognitivas y no cognitivas de los niños de cinco/seis años. *Hacienda Pública Española*, 226(3), pp. 123-153.
- MEDIIVILLA, M., MANCEBÓN, M. J., GÓMEZ-SANCHO, J. M. y JIMÉNEZ, L. P. (2019). Bilingual education and school choice: A case study of public secondary schools in the Spanish region of Madrid. *IEB Working Paper*, n.º 2019/01 Institut d'Economia de Barcelona.
- MISLEVY, R. J., BEATON, A. E., KAPLAN, B. y SHEEHAN, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), pp. 133-161.
- MURNANE, R. J. y WILLETT, J. B. (2011). *Methods matter*. Oxford University Press. New York.
- PEDRAJA, F., SANTÍN, D. y SIMANCAS, R. (2016). The impact of immigrant concentration in schools on grade retention in Spain: a difference-in-differences approach. *Applied Economics*, 48(21), pp. 1978-1990.
- PEDRAJA, F., SANTÍN, D. y SIMANCAS, R. (2022). Show me the money! The impact of a conditional cash transfer on educational achievement. *Empirical Economics*, en prensa. doi: <https://doi.org/10.1007/s00181-022-02211-x>
- PFEFFERMANN, D. y LANDSMAN, V. (2011) Are private schools better than public schools? Appraisal for Ireland by methods for observational studies. *The Annals of Applied Statistics*, 5(3), pp. 1726-1751.
- POKROPEK, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-scale Assessments in Education*, 4(1), pp. 1-20.
- SALINAS, J. y SANTÍN, D. (2012). Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en PISA 2006. *Revista de Educación*, 358, pp. 382-405.
- SANTÍN, D. y SICILIA, G. (2015). El impacto de la educación infantil en los resultados de primaria: evidencia para España a partir de un experimento natural. En: D. SANTÍN PAU BALART, A. CABRALES, J. CALERO, Á. DE LA FUENTE, J. ORIOL ESCARDÍBUL, F. FELGUEROSO, G. SICILIA (eds.), *Reflexiones sobre el Sistema Educativo Español*, pp. 45 -74. Madrid: Editorial Centro de Estudios Ramón Areces S. A.
- SCHLOTTER, M., SCHWERDT, G. y WOESSMANN, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. *Education Economics*, 19(2), pp. 109-137.
- SCHWERDT, G. y WUPPERMANN, A. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review*, 30, pp. 365-379.
- TODD, P. E. y WOLPIN, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485), pp. 3-33.
- THISTLEWAITE, D. y CAMPBELL, D. (1960). Regression-discontinuity analysis: an alternative to the ex-post facto experiment. *Journal of Educational Psychology*, 51, pp. 309-317.
- VAN DER KLAUW, W. (2008). Regression-discontinuity analysis: a survey of recent developments in economics. *Labour*, 22(2), pp. 219-245.
- VANDENBERGHE, V. y ROBIN, S. (2004). Evaluating the effectiveness of private education across countries: a comparison of methods. *Labour Economics*, 11(4), pp. 487-506.
- VILLAPLANA, C. (2014). Evaluación del programa Escuela 2.0 para la asignatura de Matemáticas a partir de PISA 2012. *Investigaciones de Economía de la Educación*, 9, pp. 631-652.
- WEBBINK, D. (2005). Causal effects in education. *Journal of Economic Surveys*, 19(4), pp. 535-560.
- WORD, E., JOHNSTON, J., BAIN, H. P., FULTON, B., ZAHARIES, J. B., LINTZ, M. N., ACHILLES, C. M., FOLGER, J. y BREDÁ, C. (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 Class Size Study: Final Summary Report*, 1985-1990. Tennessee State Department of Education, Nashville.
- WU, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2), pp. 114-128.

LA EVALUACIÓN DEL IMPACTO DE LAS POLÍTICAS PÚBLICAS: EL CASO DE LAS POLÍTICAS DEL MERCADO DE TRABAJO (*)

José M. ARRANZ
Carlos GARCÍA-SERRANO

Universidad de Alcalá

Resumen

El objetivo de este artículo es doble. Por un lado, se revisan algunos estudios que han utilizado los métodos disponibles para evaluar el impacto de las políticas públicas en el mercado de trabajo en España. Por otro lado, se lleva a cabo una aplicación empleando la técnica de dobles diferencias para evaluar un cambio legislativo que afectó a una política pasiva del mercado de trabajo español (el sistema de prestaciones por desempleo) que modificó el período de derecho de las prestaciones de los trabajadores mayores en el año 2012, para lo cual se utilizan los microdatos de perceptores del Servicio Público de Empleo Estatal.

Palabras clave: evaluación de impacto, dobles diferencias, trabajadores mayores, salidas del desempleo, prestaciones por desempleo, modelos de duración.

Abstract

The purpose of this article is twofold. On the one hand, some studies that have used the available methods to evaluate the impact of public policies on the labor market in Spain are reviewed. On the other hand, an application is carried out using the technique of double differences to evaluate a legislative change that affected a passive policy of the Spanish labor market (the unemployment benefit system) that modified the period of entitlement of the benefits of older workers in 2012, for which the microdata of recipients of the Public Employment Service are used.

Keywords: impact evaluation, double differences, older workers, exits from unemployment, unemployment benefits, duration models.

JEL classification: J64, J65.

I. INTRODUCCIÓN

UN aspecto relevante de las políticas públicas dirigidas al mercado de trabajo es saber si son efectivas o no. Por un lado, las políticas pasivas del mercado de trabajo (que básicamente comprenden el sistema de prestaciones por desempleo, SPD, tanto el nivel contributivo como el asistencial) tratan de compensar la pérdida de ingresos sufrida por las personas trabajadoras como consecuencia de la pérdida del empleo, que puede afectar negativamente a la búsqueda y la disponibilidad de las personas perceptoras. Por otro lado, las políticas activas del mercado de trabajo (la orientación y la intermediación laboral, los programas de formación y los incentivos a la contratación y el emprendimiento, principalmente) están diseñadas para compensar estos posibles efectos adversos a la vez que tratar de conseguir lo más rápidamente posible la vuelta al empleo de las personas paradas y reducir los problemas de acceso y permanencia en el empleo a los que se enfrentan ciertos grupos desfavorecidos. Para saber si estas políticas son efectivas (si realmente producen algún efecto y ayudan a las personas trabajadoras que participan en ellas) es necesario evaluarlas.

¿Por qué evaluar las políticas públicas del mercado de trabajo? La razón de evaluarlas se debe no solo a que a través de ellas es posible conocer cómo se gestionan las operaciones ejecutadas (análisis de implementación), el grado de cumplimiento de los objetivos propuestos (análisis de eficacia) y si la realización de los distintos programas generan unos beneficios mayores (o unos costes menores) que si no se hubieran llevado a la práctica (análisis de eficiencia), sino porque también permite saber si las acciones diseñadas y realizadas tienen algún efecto neto sobre ciertos indicadores relacionados con los participantes (análisis de impacto) que permitan determinar si los cambios observados se deben a la intervención de los programas y no a otros factores, y así explorar si existen estrategias alternativas con las que alcanzar mejor dichos resultados.

En este último caso, la pregunta clave es si los programas consiguen que las personas a las que se dirige la intervención mejoran su situación. Por ejemplo, en el caso particular de encontrarse desempleadas y participar en algún programa de las políticas de empleo, si dichas personas han logrado volver lo antes posible –y en las mejores

condiciones— a la actividad laboral. Solo mediante una evaluación de impacto será posible saber si estos programas realmente producen algún efecto neto en los beneficiarios y conocer si los cambios observados en ellos pueden asociarse a los efectos del programa (es decir, existe causalidad). Además, al responder a esta pregunta, se podrá comprobar si se han producido cambios estructurales como consecuencia de la actuación pública o si se han generado resultados no deseados.

Una descripción detallada de las técnicas cuantitativas de evaluación de impacto posibles puede encontrarse en otro artículo de este monográfico, así como en los múltiples manuales y referencias que existen en la literatura (véase, por ejemplo, Gertler *et al.*, 2017; Pérez y Moral-Arce, 2015 y Khandor, Koolwal y Samad, 2010, entre otros). El objetivo de este artículo no es revisar cada una de ellas, sino más bien pasar revista a cómo se han aplicado en el pasado para llevar a cabo una correcta evaluación del impacto de algunas políticas públicas vinculadas al mercado de trabajo. En concreto, el propósito del artículo es doble. Por un lado, mostrar cómo se han utilizado los métodos de evaluación de impacto en los análisis del mercado de trabajo con algunos trabajos que emplean alguno de los métodos para estimar el impacto de determinados programas en los resultados del mercado de trabajo español. Por otro lado, presentar una aplicación que analiza el impacto de un cambio legislativo en el SPD y sus efectos sobre las salidas desde el desempleo de los trabajadores mayores. Este cambio legislativo se produjo en España en julio de 2012, cuando se elevó de 52 a 55 la edad para tener derecho a una prestación asistencial (o subsidio por desempleo) ilimitada. Para ello se utilizan los microdatos de los perceptores de prestaciones por desempleo puestos a disposición por el Servicio Público de Empleo Estatal (SEPE).

El esquema del artículo es el siguiente. En la segunda sección, se ofrecen algunas aplicaciones de la literatura empírica centradas en el mercado de trabajo español que han empleado las técnicas de evaluación de impacto para cuantificar el efecto producido por algunas políticas laborales. En la sección tercera, se lleva a cabo una evaluación de impacto de una modificación normativa que afectó a una política pasiva del mercado de trabajo español. Finalmente, la última sección contiene las conclusiones.

II. REVISIÓN DE LA LITERATURA: ANÁLISIS DE EVALUACIÓN DE IMPACTO Y POLÍTICA LABORAL EN ESPAÑA

El análisis del mercado de trabajo, en general, y el examen de los efectos generados por las políticas laborales, en particular, son un campo en el que la utilización de las técnicas de evaluación de impacto resulta complicada. El problema fundamental reside en el hecho de que a menudo los cambios que se producen en los resultados laborales pueden verse influidos por una variedad de modificaciones en el entorno económico y social e incluso en las propias políticas del mercado de trabajo, cuya alteración no suele darse de manera aislada, sino que forma parte de conjuntos más amplios de cambios que afectan simultáneamente a distintos programas, como cuando los Gobiernos llevan a cabo una reforma laboral más o menos integral. Analizar el efecto de las reformas del mercado laboral es una tarea a veces desalentadora, dada la dificultad para encontrar o diseñar un grupo de control evidente o un contrafactual adecuado, sobre todo cuando se trata de medidas nuevas o cambios en los programas existentes que son uno más de muchos otros que forman parte de una reforma del mercado de trabajo (como la que se produjo en 2012). En este sentido, la interacción de una medida con otras implementadas al mismo tiempo dificulta la tarea a la hora de determinar si una medida ha contribuido a aumentar efectivamente la estabilidad en el empleo, la prevalencia de los contratos indefinidos o la probabilidad de encontrar un empleo, por ejemplo.

A pesar de ello, se ha ido creando un cuerpo (todavía bastante limitado) de estudios que han tratado de analizar el impacto de las políticas públicas en el mercado de trabajo aprovechando alteraciones en la normativa laboral y que han empleado alguna de las técnicas de evaluación de impacto existentes. En esta sección, se presenta con cierto detalle algunos de estos estudios, organizados según el diseño que llevan a cabo, haciendo hincapié en el tipo de medida que tratan de evaluar y los resultados que obtienen.

1. Diferencias en diferencias

El método de evaluación de impacto más utilizado es el de diferencias en diferencias. Dentro de las políticas activas, quizá el área en el que más se ha empleado es el del análisis de los efectos de los incentivos a la contratación y el emprendimiento. Es

el caso de los estudios que analizan el impacto que han tenido distintos programas de «bonificaciones de las cotizaciones empresariales a la Seguridad Social a la contratación de trabajadores». Algunos trabajos, centrados en el programa de incentivos a la contratación iniciado en 1997, son los de Kugler, Jimeno y Hernanz (2003) y García-Pérez y Rebollo (2009). Quizá el estudio más amplio es el de Méndez (2013), ya que analiza no solo los programas de incentivos a la contratación de 1994 y de 1997, sino también el incluido en el Plan Nacional de Empleo (PNE) para 1999 (que era una continuación del plan de 1997). Más recientemente, AIReF (2020) ha llevado a cabo una evaluación completa del sistema de incentivos a la contratación y el emprendimiento vigente en España entre 2012 y 2018, examinando por separado los incentivos a la contratación indefinida (distinguiendo entre los jóvenes, las mujeres y los trabajadores mayores), los incentivos a la conversión de contratos temporales en indefinidos, los incentivos al trabajo autónomo y los incentivos a la contratación de personas con discapacidad. La metodología que emplea en todos los casos es la de diferencias en diferencias, utilizando los datos de la *Muestra continua de vidas laborales (MCVL)* (1).

En particular, Méndez (2013) adapta los estimadores que se utilizan comúnmente en la evaluación de tratamientos específicos (es decir, los que se aplican a ciertos grupos de empleadores y/o trabajadores) a la evaluación de tratamientos no específicos (es decir, los que se aplican a todos los empleadores y trabajadores). En particular, presenta un procedimiento de estimación que utiliza períodos de tiempo de pretratamiento $T - 1$ para predecir el resultado que se habría observado en el período posterior al tratamiento en ausencia del tratamiento, siendo los estimadores antes-después, de dobles diferencias (DD) y de triples diferencias casos particulares del estimador general de T períodos. El efecto de interés se identifica utilizando el estimador más simple cuya condición de identificación no se rechaza en el período de pretratamiento.

Este autor utiliza la *Encuesta de población activa (EPA)* para el período 1987-2000 y encuentra que los empleadores reaccionaron al PNE aumentando tanto las contrataciones permanentes iniciales como las conversiones de contratos de trabajadores jóvenes y de mediana edad durante el primer semestre de 1999. Los resultados para los trabajadores de más edad fueron nulos. En particular, sus resultados apuntan a: a) efectos positivos en las tasas trimestrales de transición del desempleo al empleo

permanente para los hombres jóvenes (23 por 100) y las mujeres jóvenes (26 por 100), y los trabajadores de más edad y las mujeres de mediana edad (60 por 100), después del programa de 1997; y b) los efectos positivos en las tasas de transición trimestrales del desempleo y el empleo temporal al empleo permanente mejoraron para los trabajadores más jóvenes (35 por 100) y los trabajadores de mediana edad (71 por 100), después del PNE de 1999.

En el caso de los incentivos al emprendimiento, Cueto, Mayor y Suárez (2017) utilizan la técnica de DD para estudiar el impacto de la implantación de la «tarifa plana» para trabajadores autónomos jóvenes en el año 2013. El grupo objetivo estaba formado por varones de hasta 30 años y mujeres de hasta 35 años. Los incentivos consistieron en una reducción de la cotización mínima a la Seguridad Social para aquellos trabajadores que comenzaran un empleo por cuenta propia a partir de marzo de 2013. Los trabajadores por cuenta propia establecidos no tenían derecho a la subvención: la medida de política solo estaba disponible para las empresas de nueva creación. La reducción de las cotizaciones dependía de si la persona había tenido experiencia en el trabajo por cuenta propia durante los cinco años anteriores y de la duración del episodio de autoempleo. Si los solicitantes no habían trabajado por cuenta propia durante los cinco años anteriores, la reducción era del 80 por 100 para los primeros seis meses, del 50 por 100 para los seis meses siguientes y del 30 por 100 para los próximos 18 meses, sin distinción entre hombres y mujeres. Si los solicitantes habían trabajado por cuenta propia durante los cinco años anteriores, la reducción era del 30 por 100 durante 30 meses. El umbral de edad en ambos casos se estableció en 30 años para los hombres y en 35 años para las mujeres.

Los autores utilizan la edición de 2013 de la *MCVL* para seleccionar a las personas que entraron en el empleo por cuenta propia durante ese año e identifican a los participantes en el programa en función del cumplimiento individual de los requisitos. Luego utilizan la edición 2014 de la *MCVL* para seguir la muestra y obtener información actualizada sobre el período de trabajo por cuenta propia. Esto les permite estudiar la supervivencia en el trabajo por cuenta propia tanto de participantes como de no participantes. Sus resultados indican que las entradas en el trabajo por cuenta propia aumentaron durante los primeros meses después de la puesta en marcha de la medida de política (alrededor del 30 por 100 en los primeros cinco meses). Sin em-

bargo, utilizando la salida del trabajo por cuenta propia como variable dependiente y un modelo de duración basado en el estimador de diferencias en diferencias, el resultado es que la medida no tuvo un efecto significativo en términos de tasas de supervivencia de los nuevos autoempleados. Esto significa que la medida política no pudo ayudar a los trabajadores jóvenes a superar las desventajas potenciales a las que se enfrentan al iniciar una empresa. Tampoco encontraron efectos significativos cuando se realizaron análisis para examinar si existe algún efecto específico por género y experiencia previa de autoempleo. Parece que esta medida de política tuvo un elevado efecto de peso muerto.

El método de diferencias en diferencias también se ha utilizado para examinar programas que pueden considerarse políticas pasivas, en el sentido de que tratan de evitar la salida del empleo de los trabajadores y mantenerlos en la ocupación. Este sería el caso del «subsidio para el mantenimiento del empleo de los trabajadores mayores», que estuvo en vigor en el período 2006-2012 y dirigida a los trabajadores de 60 a 64 años, ya contratados con un contrato indefinido y con al menos cinco años de antigüedad dentro de una empresa. Esta medida no tenía por objeto aumentar la contratación, sino mantener en el empleo a un grupo definido de trabajadores (en riesgo de desempleo y/o de abandonar la población activa). Font, Izquierdo y Puente (2021) evaluaron el efecto que este subsidio pudo tener sobre la probabilidad de que los trabajadores de edad avanzada permanecieran en la misma empresa a partir de los datos de la *MCVL*. Los autores estiman la probabilidad de perder sus empleos entre agosto y finales de año para los trabajadores mayores que mantuvieron su empleo hasta julio de 2012. Consideran que la eliminación de las subvenciones al mantenimiento del empleo en julio de 2012 es el tratamiento, y los trabajadores de 60 años o más sujetos a un contrato indefinido y con al menos cinco años de antigüedad dentro de la empresa como el grupo tratado. Para inferir el resultado contrafactual, explotan la continuación de los subsidios para el mantenimiento del empleo de los trabajadores de 59 años hasta diciembre de 2012; de ahí que su grupo de control consista en aquellos trabajadores de 59 años con contrato indefinido y con al menos cinco años de antigüedad en la empresa.

Sus resultados indican que la diferencia en la probabilidad de separación de la empresa en el grupo tratado (60 a 64 años) frente al grupo de control (59 años) aumentó 1,8 puntos porcentuales

(p. p.) después de que se suprimieran los subsidios. Además, aunque este aumento se produjo para todos los niveles de edad (1,4 p. p. para los trabajadores de 60 a 63 años), fue mayor y estadísticamente significativo para los trabajadores de 64 años (un aumento de 2,8 p. p.). Los autores realizan análisis adicionales por antigüedad y nivel de cualificación, encontrando que los efectos positivos se concentran en los trabajadores con menor antigüedad (5-9 años) y en los puestos de trabajo con menores requisitos de cualificaciones (aunque en este caso las diferencias no son estadísticamente significativas).

Estos resultados apoyan el argumento de que los incentivos al mantenimiento del empleo para los trabajadores de edad madura se asociaron positivamente con una mayor retención del empleo mientras el subsidio estaba en vigor. Cuando se eliminaron los incentivos, los trabajadores que anteriormente habían tenido derecho a una reducción de sus cotizaciones a la Seguridad Social vieron crecer la probabilidad de perder su empleo. Además, el subsidio de mantenimiento del empleo habría contribuido a generar más estabilidad laboral y menos transiciones hacia la desocupación para aquellos trabajadores con mayor riesgo de ser excluidos permanentemente del mercado laboral.

Finalmente, el método de DD también se ha empleado para tratar de evaluar otras medidas relacionadas con el mercado de trabajo, como el impacto de las subidas del salario mínimo interprofesional (SMI). Es el caso de Galán y Puente (2015), que se centran en los incrementos llevados a cabo durante la primera década de los años 2000. La metodología que utilizan es la que se ha empleado en otros estudios llevados a cabo por el Banco de España para analizar los aumentos del SMI que se han producido posteriormente: el del 2017, un 8 por 100 en términos nominales (Banco de España, 2017) y el de 2019, un 22,3 por 100 en términos nominales (Lacuesta, Izquierdo y Puente, 2019). Galán y Puente (2015) examinan los efectos de las subidas que se hicieron efectivas entre 2004 y 2010, generando un incremento acumulado de un 37,5 por 100 en términos nominales (18,7 por 100 en términos reales). Estos aumentos implicaron un crecimiento sostenido del SMI real por primera vez desde finales de la década de 1970, dando lugar a que en 2008 hubiese recuperado el poder adquisitivo que tenía en 1980.

Los autores analizan el período 2000-2010, que incluye un subperíodo sin aumentos del SMI

real (2000-2004) y otro con aumentos del SMI real (2005-2009), utilizando el primero como control para el efecto de los salarios bajos sobre la probabilidad de empleo. La estrategia de identificación consiste en comparar trabajadores con un salario real por debajo del próximo mínimo durante el segundo período con trabajadores de similares características que tienen un salario real por encima (pero cerca) del mínimo, que por tanto no se ven afectados por el aumento, y también con trabajadores con similares características y mismo salario durante el primer período. Utilizando datos administrativos procedentes de la *MCVL* y definiendo la variable dependiente como la probabilidad de perder un empleo doce meses después, el resultado que obtienen es que las subidas del SMI fueron negativas para el empleo: un aumento en el SMI induce un impacto significativamente positivo en las tasas de salida del empleo para los trabajadores jóvenes (de 16 a 24 años) y mayores (de 46 a 64 años). Estas transiciones fuera del empleo corresponden a un aumento de la probabilidad de quedar desempleado para ambos grupos y también a un impacto positivo en la probabilidad de convertirse en autónomo para los trabajadores de más edad.

2. Emparejamiento

Otros estudios que examinan las políticas laborales han empleado el método de emparejamiento. Es el caso de Jansen y Troncoso (2018) y Cueto y López (2019), que se centran en un cambio producido en el «contrato de formación y aprendizaje» (CFA). Este tipo de contrato está dirigido a jóvenes de entre 16 y 24 años con solo educación primaria o secundaria. Su objetivo es contribuir a incrementar la cualificación profesional de estos trabajadores alternando la formación con la actividad laboral en una empresa, y mejorar su integración en el mercado laboral. Esta medida de política se considera un instrumento para luchar contra el desempleo juvenil y la segmentación del mercado laboral y se ha utilizado desde la década de 1990, sufriendo varias reformas desde el comienzo de la crisis financiera en 2008, con modificaciones llevadas a cabo por distintos Gobiernos en el período 2010-2012, introduciendo cambios que pretendían hacerlo más atractivo para los empresarios.

Cueto y López (2019) centran su análisis de evaluación en la reforma de 2011 utilizando datos procedentes de la *MCVL* referidos al período 2011-2015. Entre otros cambios, la reforma estableció

nuevas normas sobre la duración de los CFA, aumentando la duración mínima a un año; y elevó el límite de edad para acceder a dichos contratos de 25 a 29 años (esto fue motivado por el fuerte aumento observado en el número de jóvenes «ninis» con bajo nivel educativo y de edad superior a 25 años). Este hecho ofrece un escenario único para evaluar el impacto de los CFA en la integración de los jóvenes en el mercado laboral español ya que, antes de la implantación de esta medida, este colectivo de jóvenes de entre 25 y 29 años podía no acceder a un dicho tipo de contrato.

Aprovechando este escenario para evaluar la efectividad de los CFA, los autores crean un grupo de tratamiento y otro de control. El primero está formado por el conjunto de contratos de formación y aprendizaje a tiempo completo firmados en 2011, mientras que el segundo incluye los contratos temporales por obra o servicio y eventuales por circunstancias de la producción. Del grupo de control se excluyen a las personas que han tenido contratos de formación en el pasado. Además, como la entrada en vigor de la reforma fue el 31 de agosto de 2011, la muestra se divide en dos: un grupo prerreforma (que incluye todos los contratos firmados antes de la introducción de la reforma –de enero a agosto de 2011–) y un grupo posreforma (que incluye todos los contratos firmados después de la introducción de la reforma (de septiembre a diciembre de 2011)). La muestra se limita a los jóvenes de entre 16 y 25 años, ya que este fue el límite de edad aplicable durante el primer semestre de 2011. En cuanto al grupo de control, solo se seleccionan aquellos contratos temporales que tuvieron una duración superior a seis meses para que ambos grupos sean lo más homogéneos posible. Dado que se eliminan los contratos temporales de muy corta duración, el estudio analiza si el contrato de formación, con sus propias peculiaridades, contribuye a una mejor inserción laboral que un contrato temporal de la misma duración.

Los autores utilizan varias variables relacionadas con los resultados laborales: si la persona trabajó en algún momento una vez finalizado el contrato; si obtuvo acceso a un contrato permanente; el tiempo trabajado como porcentaje del tiempo de trabajo potencial al final del contrato; y los salarios mensuales promedio en los años siguientes después de la contratación. Con casi todas las variables de resultado analizadas, se obtienen efectos significativamente negativos. Ser contratado bajo un CFA con una duración mínima de seis meses para un joven entre 16 y 25 años es menos efectivo que ser contratado

bajo un contrato temporal similar, tanto en términos de probabilidad de ser contratado para otro trabajo (casi un 30 por 100 menos), del porcentaje de tiempo trabajado (aproximadamente un 30 por 100 menos) y de la remuneración al final del contrato (alrededor de 400 euros mensuales menos después de un año; alrededor de 200 euros mensuales menos después de tres años). Por lo que se refiere al acceso a un contrato indefinido, el efecto del tratamiento no es significativo, por lo que indica que el CFA no muestra ni ventajas ni desventajas con respecto a un contrato temporal (este efecto es ligeramente positivo solo cuando se consideran contratos más largos –un año o más–). Por tanto, se puede decir que los CFA no parecen conducir a mejores resultados en el mercado laboral de los jóvenes. El único aspecto positivo es que los CFA tienen una duración más larga (3-4 meses más) que los temporales. Esto puede estar relacionado con el hecho de que la normativa legal impide que este tipo de contratos se encadenen en el tiempo (aunque esta protección se redujo con la reforma de 2012). Jansen y Troncoso (2018) también analizan la reforma aprobada en 2011 con una metodología similar y llegan a unos resultados parecidos: el aumento temporal del límite de edad para acceder a un CFA no produjo una mejora significativa en la tasa de empleo de las personas afectadas por la medida.

Un trabajo reciente que también utiliza el método de emparejamiento es el de Rebollo y García-Pérez (2021). En su caso, los autores realizan un ejercicio de evaluación de impacto de un tipo de programa que se dirige a parados de difícil inserción laboral (las iniciativas locales de empleo) y, en particular, dos medidas desarrolladas en Andalucía en 2016-2018 («Emple@Joven», dirigido a parados menores de 30 años, y «Emple@30+», dirigido a parados mayores de 30 años). Utilizan datos procedentes de los ficheros de demandantes de empleo del Servicio Andaluz de Empleo (SAE) para delimitar los grupos de tratamiento y de control de cada medida, así como información sobre la situación laboral de beneficiarios y no beneficiarios seis meses después de haber finalizado el contrato asociado a los programas y, en caso de estar ocupados, sobre las características del empleo. Los resultados del estudio apuntan a que el programa «Emple@Joven» no aumenta los niveles de empleabilidad para los beneficiarios del programa en comparación con no beneficiarios similares, mientras que el programa «Emple@30+» solo favorece la empleabilidad de los parados sin experiencia laboral o aquellos que han estado fuera del mercado de trabajo durante más de veinticuatro

meses. Sin embargo, para aquellos que están ocupados posteriormente, las características del empleo de los tratados son algo mejores que las de los controles (mayor probabilidad de estar ocupado a tiempo completo y menor probabilidad de que los contratos sean de corta duración y/o de baja cualificación).

3. Regresión en discontinuidad

En cuanto al método de regresión en discontinuidad, los estudios son más escasos. Un caso interesante lo constituye el estudio de Gamberoni, Gradeva y Weber (2016) que analiza el impacto del «contrato de apoyo a emprendedores» utilizando datos individuales de empresa (provenientes de «Amadeus», una base de datos comercial en el ámbito internacional que se basa en la información de los balances de las empresas). Este tipo de contrato indefinido se introdujo en febrero de 2012, para trabajadores jóvenes y previamente desempleados contratados por empresas pequeñas, con el objetivo de luchar contra el desempleo y la rotación excesiva. Este contrato estaba dirigido a los trabajadores desempleados desfavorecidos (jóvenes de 16 a 30 años y trabajadores mayores de 45 años registrados en el SEPE) y a las empresas pequeñas (empresas que tenían menos de 50 empleados antes de hacer uso del contrato). Se podía utilizar tanto para el empleo a tiempo completo como para el empleo a tiempo parcial, implicaba un período de prueba ampliado de doce meses (con la posibilidad de terminar el contrato a voluntad durante ese tiempo) e incorporaba varios incentivos financieros (aplicables hasta que la tasa global de desempleo fuese inferior al 15 por 100). Los incentivos incluían importantes desgravaciones fiscales y reducciones en las cotizaciones patronales a la Seguridad Social, que variaban según la edad y el período de desempleo anterior del trabajador contratado, incluida una deducción fiscal del 50 por 100 de las prestaciones por desempleo que recibiría el trabajador desempleado en el momento de su contratación con el nuevo contrato permanente. Los incentivos estaban condicionados a mantener al trabajador al menos tres años dentro de la empresa.

Como la subvención solamente se concedió a empresas que tenían menos de 50 empleados, el límite de tamaño arbitrario proporciona un candidato natural para analizar las respuestas de las empresas comparando las empresas tratadas con las empresas de control, con características similares, en el entorno del umbral de 50 empleados. Gamberoni, Gradeva y Weber (2016) encuentran pequeños efec-

tos positivos de la medida de política a corto plazo: el crecimiento del empleo de las empresas elegibles es aproximadamente un 2 por 100 más alto que el crecimiento del empleo de las empresas que no podrían beneficiarse de la subvención. Sin embargo, las pruebas complementarias (basadas en regresiones de diferencias en discontinuidades y excluyendo a las empresas que potencialmente se clasificaban deliberadamente por debajo del umbral) hacen que los efectos estimados disminuyan, de modo que la diferencia entre el crecimiento del empleo de las empresas elegibles y las empresas de control ya no es estadísticamente significativa. Aplicando los criterios de umbral a los cuatro años anteriores a la reforma, los autores encuentran la presencia de posibles factores de «selección» en el año anterior a la reforma (2011). Tener en cuenta este aspecto con el uso del enfoque de diferencias en discontinuidades y otras pruebas de robustez implica que la subvención no tuvo un efecto significativo en el crecimiento del empleo. Por tanto, el crecimiento diferencial de las empresas en torno al umbral de los 50 empleados parece derivarse más bien del menor crecimiento después de 2011 de las empresas de más de 50 empleados que desean eludir los requisitos contables más estrictos como resultado de las nuevas regulaciones que imponían requisitos de información más estrictos a las empresas más grandes aprobadas en el contexto de la reforma de 2010.

III. UNA APLICACIÓN: EVALUACIÓN DE UNA POLÍTICA LABORAL EN ESPAÑA

Debido al elevado y persistente desempleo durante la década de 1980, muchos países europeos idearon medidas para reducir la oferta efectiva de mano de obra, extendiendo los períodos de derecho de las prestaciones por desempleo para los trabajadores mayores y/o adaptando planes de jubilación anticipada para ellos, de modo que las prestaciones y subsidios por desempleo proporcionaron un camino para su retirada anticipada del mercado laboral (Duval, 2003). Estos programas generaban fuertes incentivos para que los trabajadores mayores despedidos antes de tener derecho a una pensión completa recibieran prestaciones por desempleo y utilizaran el sistema de prestaciones como vía hacia la jubilación. Al mismo tiempo, desde el punto de vista de las empresas, despedir primero a los trabajadores mayores podía parecer más aceptable socialmente que colocar la carga de la pérdida del empleo en otros empleados.

Los programas de jubilación anticipada a menudo presentaban períodos prolongados de la duración máxima de las prestaciones contributivas por encima de ciertos límites de edad (Schmieder, Von Wachter y Bender, 2016) o criterios de elegibilidad para recibir las prestaciones por discapacidad relajados para los trabajadores mayores (Chen y Van der Klaauw, 2008). El sistema español de jubilación anticipada basado en las prestaciones por desempleo era similar al de otros países europeos (Belloni, Monticone y Trucchi, 2006). En España, existía el denominado «subsidio para personas mayores de 52 años», mediante el cual las personas desempleadas que cumplieran 52 años después de agotar su prestación contributiva (PC), habiendo cotizado durante al menos seis años durante su vida laboral y cumpliendo todas las condiciones (excepto la edad) para recibir una pensión de la Seguridad Social, podían cobrar una prestación asistencial (PA) hasta el momento en que fuesen elegibles para una pensión de jubilación normal. Estos términos cambiaron en 2012, cuando la edad para ser elegible para este subsidio se elevó de 52 a 55 (entonces, el subsidio tomó el nombre de «subsidio para personas mayores de 55 años»). Esta reforma redujo el período de derecho ilimitado para el grupo de personas de 52 a 54 años, proporcionando un escenario ideal para una evaluación cuasiexperimental de la influencia de la duración potencial de una prestación en la duración efectiva del desempleo. Aprovechando esta reforma, y utilizando dos dimensiones disponibles para la identificación (la elegibilidad –edad– y el tiempo –antes y después del cambio normativo–), podemos identificar los efectos de las PA sobre la duración del desempleo (2).

1. El contexto: el sistema de prestaciones y la reforma de 2012

El SPD español dispone de dos tipos de prestaciones: PC y PA. Las PC se pagan a los empleados (excluidos los funcionarios públicos, los trabajadores contratados por hogares y aquellos sin experiencia laboral previa) que perdieron su trabajo o cuyo contrato temporal terminó, que pueden y quieren trabajar y que han cotizado un tiempo mínimo en el empleo (al menos doce meses durante el período de referencia de los seis años anteriores). Las PC se pueden recibir por un mínimo de cuatro meses y un máximo de veinticuatro meses, dependiendo del tiempo trabajado (una persona tiene derecho al máximo si ha trabajado durante todo el período de referencia). La cantidad de PC pagada es igual a

una fracción del salario bruto de los seis meses anteriores al desempleo y disminuye con la duración del período: la tasa de reemplazo bruta es del 70 por 100 durante los meses 1 a 6 de la recepción de PC y del 60 por 100 a partir de entonces.

Los trabajadores que no son elegibles para una PC (porque no han acumulado el período mínimo de contribución) pueden acceder a una PA. Aquellos que han agotado la PC también pueden solicitar una PA. La percepción de una PA requiere ciertas condiciones de elegibilidad: estar registrado durante al menos un mes en alguna oficina pública de empleo; no haber rechazado ninguna oferta de empleo adecuada o participación en un programa de formación o cualificación profesional dirigido por el SEPE; y tener ingresos inferiores al 75 por 100 del salario mínimo. La cuantía de la PA es fija, el 80 por 100 del indicador público de renta de efectos múltiples (IPREM), es decir, 452 euros en 2021. La duración del período de derecho de las PA es de al menos seis meses y varía según el número de familiares a cargo y la edad del beneficiario (el máximo es de 36 meses).

Las personas mayores que pierden el empleo no están, en principio, sujetas a PC más generosas. Sin embargo, existe una excepción en el sistema de PA anterior. Los desempleados que cumplieron 52 años durante su período de derecho potencial y cotizaron a la Seguridad Social por el riesgo de desempleo durante al menos seis años durante su vida laboral y cumpliendo todas las condiciones (excepto la edad) para recibir una pensión de la Seguridad Social, podían reclamar el «subsido para mayores de 52 años» que forma parte de las PA. Las condiciones que debían cumplir para recibir este tipo de subsidio eran las mismas que las generales de las PA («Otros tipos de PA»). Recibir el «subsido para mayores de 52 años» permitía a esos trabajadores percibir prestaciones por desempleo hasta el momento en que eran elegibles para una pensión de jubilación normal, siendo un subsidio con un período de derecho de la PA ilimitado. Ambas características estaban ausentes en el resto de las PA. Así, las PC y PA regulares seguidas de este subsidio especial para trabajadores mayores potencialmente les permitían retirarse del mercado laboral a una edad muy temprana.

Esta situación cambió en julio de 2012 cuando en la reforma del mercado laboral aprobada por el nuevo Gobierno español, la edad para acceder a este subsidio especial se elevó de 52 a 55 años. Por tanto, a partir de julio de 2012 una persona desempleada debe de tener al menos 55 años cuando agota una

PC para tener la opción de cobrar el «subsido para personas mayores de 55» hasta el ingreso posterior al régimen de pensiones. La reforma se aplicó con derechos adquiridos, de modo que todos los beneficiarios que estaban percibiendo un «subsido de 52 años» seguirían recibéndolo incluso si el requisito de edad ya no se cumpliera después de que se aprobara la reforma. Este cambio se adoptó en el contexto de una reforma del mercado laboral más general. Como no hay razón para creer que el umbral de edad se elevó en respuesta a un cambio en las condiciones relativas del mercado laboral para los trabajadores mayores, nuestro análisis no debería estar sujeto a sesgos de política endógenos.

2. Identificación

La reforma de 2012 redujo efectivamente el período de derecho a prestaciones entre el grupo de 52 a 54 años, proporcionando un entorno cuasiexperimental para estudiar la relación entre la duración de las PA y las transiciones para salir del desempleo. A partir de los datos del SEPE, extraeremos dos muestras: una muestra «antes de la reforma» y una muestra «después de la reforma». En nuestro primer escenario, los trabajadores de 49 a 51 años que percibieron una PA porque agotaron sus PC en 2011 y 2013, respectivamente, serán etiquetados como grupos de control antes y después de la reforma. Ninguno de los dos fue elegible para recibir el «subsido de 52/55 años». Los trabajadores de 52 a 54 años que percibieron la PA en 2011 son considerados el grupo de tratamiento antes de la reforma, puesto que su comportamiento se vio potencialmente afectado por su elegibilidad para el «subsido de 52/55 años». El grupo de tratamiento posterior a la reforma estaría compuesto por trabajadores de 52 a 54 años que percibieron una PA después del agotamiento de la PC en 2013, ya que ya no serían elegibles para percibir el «subsido de 52/55 años» en esas edades. En nuestro segundo escenario, los trabajadores de 55 a 57 años que percibieron una PA debido al agotamiento de las PC en 2011 y 2013, respectivamente, se consideran los grupos de control antes y después de la reforma. En este caso, ambos fueron elegibles para el «subsido de 52/55 años».

Para estimar el efecto de las prestaciones por desempleo en los resultados de los trabajadores, adoptaremos un método de DD para comparar las transiciones desde el desempleo antes y después de la reforma para los grupos afectados (tratados) y no afectados (control) por la reforma. Que el estimador DD identifique correctamente el efecto causal de la

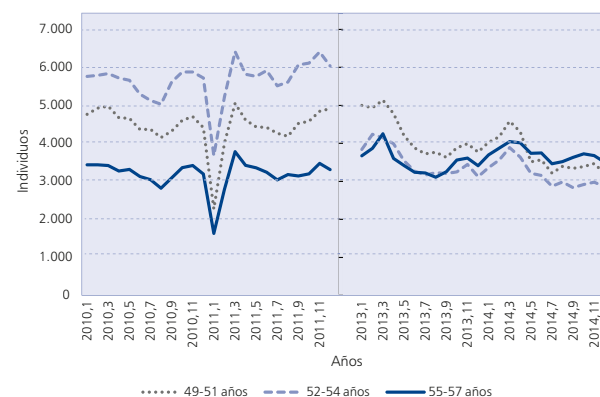
duración extendida del subsidio sobre el tránsito desde el desempleo depende de si el cambio de política fue exógeno. Hay dos razones por las que la endogeneidad de esta política es probablemente de menor importancia en nuestro contexto. Por un lado, la presión política para tratar a los trabajadores en riesgo de despido permanente y a los trabajadores desempleados que aún están en su mejor momento (menores de 55 años) con menos generosidad que antes fue una de las razones para cambiar las reglas del acceso a las prestaciones por desempleo. En la medida en que tales preocupaciones de equidad fueron la razón del cambio de política, la reducción en los períodos de derecho potenciales de las prestaciones (a través de la eliminación de la extensión de las PA para las personas de 52 a 54 años) puede considerarse exógena con respecto a los resultados del mercado laboral de las personas elegibles en nuestra muestra.

Por otro lado, la segunda razón es que la economía española estaba en recesión económica antes del cambio de política (en los años 2010 y 2011). Después del cambio de política (en los años 2013 y 2014), la economía española estaba iniciando una rápida recuperación. La endogeneidad de la política no es un problema en este caso porque todos los grupos de edad considerados en nuestro análisis estaban sufriendo los efectos de la recesión de una manera similar antes del cambio de política y beneficiándose de la recuperación después. Esta característica está relacionada con el supuesto de «tendencia común» en el que se basa la estrategia DD. En nuestro análisis utilizamos el grupo 49-51 (alternativamente, el grupo 55-57) como grupo de control para examinar los cambios en las entradas en el desempleo en ausencia de la reducción de la duración de los derechos. El gráfico 1 ofrece la información sobre las tendencias en las entradas de PA durante el período 2010-2014 y proporciona evidencia de que las entradas de los grupos de edad 49-51 y 52-54 (y también de 55-57) se movieron de manera similar en 2010-2011 y no mostraron diferencias sustanciales en la tendencia durante el período anterior a la reforma.

3. Base de datos y descriptivos

La base de datos utilizada en este estudio procede de los registros administrativos del SEPE español. En estos registros administrativos las observaciones de cada individuo contienen información sociodemográfica de los demandantes de empleo, incluidas las características del trabajador (género, edad, nacio-

GRÁFICO 1
FLUJO DE ALTAS DE SUBSIDIOS POR DESEMPLEO PARA LOS TRES GRUPOS DE EDAD (49-51, 52-54 Y 55-57)



Fuente: Ficheros del SEPE (enero 2010-diciembre 2014).

alidad y municipio y provincia de residencia) y los atributos del último empleo (ocupación y sector), así como la causa de salida del empleo (despido individual, fin de contrato temporal, etc.). La base de datos también contiene información sobre el tipo de prestaciones por desempleo, la cuantía y los períodos de derecho de las prestaciones en el momento de la entrada y la duración efectiva del desempleo.

Para el análisis utilizaremos todos los trabajadores de 49 a 57 años que comenzaron a recibir PA en 2011 (período previo a la reforma) y en 2013 (posterior a la reforma). Podemos seguirlos hasta septiembre de 2017 (último mes disponible). Para cada trabajador de la muestra, observamos la duración del período de desempleo (en días), su tipo de salida y un conjunto de variables de control. Hacemos una distinción entre salidas desde PA hacia el empleo y «Otras PA». Todos los períodos de desempleo que continúan más allá de los tres años desde el momento en que ingresaron al desempleo se tratarán como períodos censurados artificialmente en los modelos estimados para evaluar el impacto de la reforma, porque los beneficiarios del «subsidio de 52/55 años» pueden recibir este tipo de prestaciones hasta la jubilación.

La credibilidad de nuestra identificación depende de la similitud entre los grupos de tratamiento y control. El cuadro n.º 1 proporciona estadísticos descriptivos para los grupos de edad de tratamiento y control antes y después del cambio de política,

CUADRO N.º 1

MEDIAS DESCRIPTIVAS DE LAS ENTRADAS A UN SUBSIDIO POR DESEMPLEO: GRUPOS DE CONTROL Y TRATAMIENTO, ANTES Y DESPUÉS DEL CAMBIO LEGISLATIVO (PERCEPTORES DE SUBSIDIOS: GRUPOS DE EDAD 49-51, 52-54 Y 55-57)

	AÑO 2011 (ANTES DEL CAMBIO LEGISLATIVO)			AÑO 2013 (DESPUÉS DEL CAMBIO LEGISLATIVO)			DIFERENCIAS DESPUÉS-ANTES		
	CONTROL I 49-51	TRATADO 52-54	CONTROL II 55-57	CONTROL I 49-51	TRATADO 52-54	CONTROL II 55-57	CONTROL I 49-51	TRATADO 52-54	CONTROL II 55-57
Edad	50,5	53,1	56,5	50,5	53,4	56,3	0,0	0,4	-0,2
Sexo (varones)	0,631	0,629	0,652	0,611	0,619	0,653	-0,020	-0,010	0,002
Nacionalidad (española)	0,893	0,928	0,913	0,913	0,917	0,932	0,020	-0,011	0,019
Sector									
Valores perdidos	0,010	0,021	0,019	0,010	0,009	0,024	0,000	-0,012	0,005
Agricultura y pesca	0,047	0,045	0,043	0,050	0,045	0,044	0,003	-0,001	0,001
Industria	0,156	0,179	0,185	0,125	0,145	0,166	-0,030	-0,035	-0,019
Construcción	0,234	0,207	0,229	0,191	0,200	0,201	-0,042	-0,007	-0,028
Servicios	0,554	0,547	0,524	0,624	0,601	0,565	0,070	0,054	0,041
Ocupaciones									
Directores y gerentes	0,008	0,009	0,011	0,010	0,010	0,011	0,002	0,002	0,000
Profesionales, científicos e intelectuales	0,025	0,031	0,025	0,046	0,040	0,034	0,020	0,008	0,009
Técnicos profesionales de nivel medio	0,045	0,055	0,061	0,067	0,068	0,070	0,022	0,013	0,009
Personal de apoyo administrativo	0,072	0,091	0,091	0,102	0,098	0,098	0,030	0,007	0,007
Trabajadores de servicios	0,128	0,134	0,125	0,181	0,179	0,168	0,053	0,045	0,043
Agricultores y trabajadores cualificados	0,015	0,016	0,016	0,017	0,016	0,016	0,001	0,000	0,000
Oficiales, operadores y artesanos	0,213	0,224	0,238	0,217	0,227	0,250	0,004	0,003	0,012
Operadores de instalaciones y máquinas ensambladoras	0,082	0,101	0,090	0,097	0,099	0,099	0,014	-0,001	0,009
Ocupaciones elementales	0,283	0,317	0,311	0,264	0,263	0,256	-0,018	-0,054	-0,055
Militares	0,006	0,003	0,004	0,000	0,000	0,000	-0,006	-0,003	-0,004
Valores perdidos	0,121	0,020	0,029	0,000	0,000	0,000	-0,121	-0,020	-0,029
Número de hijos	0,832	0,230	0,196	0,865	0,628	0,331	0,033	0,398	0,135
Causa de entrada a PA									
Otras causas (agotamiento)	1,000	0,206	0,241	1,000	1,000	0,301	0,000	0,794	0,060
Subsidio 52/55 años	0,000	0,794	0,759	0,000	0,000	0,699	0,000	-0,794	-0,060
Causa de salida de PA									
Otras PA	0,225	0,115	0,131	0,226	0,321	0,175	0,001	0,206	0,043
Empleo	0,224	0,161	0,159	0,354	0,289	0,234	0,129	0,128	0,075
Agotamiento (censura)	0,551	0,724	0,454	0,420	0,389	0,591	-0,131	-0,334	0,137
Período de derecho	478,9	3456,6	2397,3	581,4	535,6	1775,3	102,5	-2921,0	-621,9
Duración efectiva	334,3	802,9	758,1	372,0	309,9	645,9	37,7	-493,0	-112,2
Nivel PA (€/día)	14,7	14,4	14,4	13,1	13,2	13,4	-1,6	-1,2	-0,9
Salarios (€/mes)	532,5	532,5	532,5	532,5	532,5	532,5	0,0	0,0	0,0
Número de individuos	52.141	68.578	37.585	50.716	42.473	42.243			

Fuente: Datos del SEPE (2010-2017).

así como la diferencia «después-antes» para cada grupo. Como puede verse, las características sociodemográficas y laborales parecen estar bastante equilibradas entre las muestras.

Este cuadro también ofrece información sobre las entradas y las salidas de los grupos de tratamiento y de control. En cuanto a las entradas, la proporción de perceptores tratados que entran en el «subsidio de 52/55 años» después de agotar la PC se redujo mucho (79 por 100) en 2013 en comparación con 2011, lo que sugiere que la reforma redujo efectivamente las transiciones de PC al subsidio para trabajadores mayores. Aunque este efecto es grande, se debe en gran parte a la aplicación de la ley. Al mismo tiempo, aumentó la proporción de entradas en otras PA (debido al agotamiento de la PC), lo que simplemente refleja la sustitución del «subsidio de 52/55 años» por «Otras PA» para las personas que ya no eran elegibles para el subsidio ilimitado hasta cumplir los 55 años.

Por lo que respecta a las salidas, una gran parte de los períodos de desempleo del grupo de tratamiento antes de la reforma duraron más allá del período de observación, como lo indica la tasa de censura (72 frente a 38,9 por 100). De hecho, las salidas al empleo se observaron con mayor frecuencia en el grupo de control más joven que en el grupo tratado tanto antes del cambio de política (22,4 frente al 16,1 por 100) como después del cambio de política (35,4 frente al 28,9 por 100). Al mismo tiempo, las tasas de transición al empleo del grupo de control de mayor edad fueron cercanas a las del grupo de tratamiento tanto antes (15,9 por 100) como después del cambio de política (23,4 por 100).

4. Modelo econométrico

En nuestro caso, las tasas de salida de las PA se analizarán utilizando un modelo de duración de riesgo proporcional mixto de tipo continuo (*mixed proportional hazard model*). La representación de la tasa de salida del modelo es la siguiente:

$$h_i(t_i / X, \theta) = \lambda_{0i}(t_i) \varphi_i(X'\beta_i) \phi_i(\theta) \quad i=1,2,\dots,N \quad [1]$$

donde t_i es la duración actual del período de PA para cada individuo i . Esta ecuación muestra que la tasa de salida del desempleo bajo PA hacia el empleo se ve afectada por tres componentes. Primero, la función $\lambda_{0i}(t)$, denominada función de riesgo básico, captura el efecto del tiempo transcurrido en

el estado de desempleo bajo PA sobre la probabilidad instantánea de encontrar un trabajo cuando el resto de los factores o componentes se mantienen constantes. Segundo, la función $\varphi_i(X)$ representa el efecto de las variables explicativas en la probabilidad de encontrar un empleo. Y tercero, la función $\Phi_i(\theta)$ recoge los efectos del componente de la heterogeneidad inobservada, como es la capacidad, las actitudes, las habilidades, etc., de los trabajadores sobre la probabilidad de encontrar un trabajo. Las tres funciones deben asegurar que la expresión [1] no sea negativa. Garantizamos esta propiedad mediante el uso de una representación exponencial para cada función. Para capturar el efecto de las variables explicativas sobre la tasa de riesgo de encontrar un empleo, se utiliza la siguiente especificación:

$$\varphi_i(X'\beta) = [\exp(\beta_0 + \beta_1 * Z_i + \beta_2 * R_i + \beta_3 * Age52_54_i + \beta_4 * R_i * Age52_54_i)] \quad [2]$$

donde X_i se puede dividir en varias partes: Z_i representa un vector de variables de control exógenas; R_i es una variable ficticia para el cambio de política adoptado en 2012 (toma el valor 0 para personas que comenzaron un período de PA en 2011, es decir, antes de que tuviera lugar el cambio de política; y el valor 1 para aquellas que comenzaron un período de PA en 2013, después del cambio de política); y $Age52_54_i$ es una variable ficticia que indica si el individuo pertenece al grupo de control (=0) o al grupo de tratamiento (=1).

Por tanto, β_2 mide el efecto del momento del tiempo (calendario) en las transiciones desde las PA al empleo que es independiente del estado de elegibilidad de la observación i ; β_3 es la probabilidad media de transición de los individuos pertenecientes al grupo de edad de 52 a 54 años; y β_4 el correspondiente estimador de DD que captura el impacto de tener entre 52 y 54 años después del cambio de política, es decir, cuando se eliminó la elegibilidad a la extensión del subsidio para ese grupo de edad. Este último parámetro identifica el efecto causal bajo el supuesto de «tendencia común», es decir, que cualquier factor no observable variable en el tiempo tiene el mismo efecto en los trabajadores tratados y no tratados. La interpretación de este parámetro es que mide la mayor o menor probabilidad de salir del desempleo en el período posterior al cambio de 2012, en relación con los beneficiarios no tratados y en relación con el período anterior al cambio legislativo.

Para estimar el modelo de duración, se puede considerar un modelo paramétrico que hace fuertes supuestos sobre la forma de la función de riesgo básico o un modelo de Cox que no hace ninguna. Nosotros, elegimos un modelo de riesgo semiparamétrico exponencial por partes (*piecewise constant exponential model*) que se encuentra entre ambos, que es muy utilizado en modelos de duración de tiempo continuo. En este tipo de modelos, se supone que el riesgo de salida es constante dentro de los intervalos de tiempo de supervivencia prespecificados, pero esta salida constante puede diferir entre intervalos (Jenkins, 2005). En nuestro caso, permitimos que el riesgo difiera en intervalos de treinta y seis meses y usamos la distribución gamma (una opción popular debido a la trazabilidad matemática) para la distribución de la heterogeneidad individual inobservada.

5. Resultados de las estimaciones

En esta subsección, presentamos los resultados empíricos de las estimaciones del modelo presentado en la subsección anterior. En estas estimaciones, la variable dependiente es la duración de las PA (en

días) y el conjunto de variables de control son el sexo, la nacionalidad, el número de hijos, la ocupación, el sector de actividad económica y la tasa de desempleo regional. La columna 1 del cuadro n.º 2 proporciona los resultados de las estimaciones del modelo semiparamétrico de duración con heterogeneidad inobservada para la muestra total, utilizando los perceptores de PA de 49 a 51 años antes y después del cambio de política como grupo de control (grupo de control I).

Los parámetros de los modelos de duración estimados se muestran en términos de la métrica del modelo de riesgo proporcional (*proportional hazard, PH*) en lugar de la métrica del tiempo de riesgo de salida acelerado (*accelerated failure time, AFT*). Esto significa, por ejemplo, que un coeficiente positivo implica un aumento en la tasa de riesgo (en términos logarítmicos), mientras que un coeficiente con signo negativo implica una disminución. Hemos estimado los modelos sin y con heterogeneidad inobservada (utilizando una distribución gamma). La prueba de razón de verosimilitud de un modelo con heterogeneidad inobservada frente a otro sin ella sugiere que la heterogeneidad inobservada no

CUADRO N.º 2

RESULTADOS DE LAS ESTIMACIONES DEL MODELO DE DURACIÓN SEMIPARAMÉTRICO DE LA ECUACIÓN [2]: CON EL GRUPO DE CONTROL I, PERCEPTORES DE EDADES 49-51 (COLUMNA 1); CON EL GRUPO DE CONTROL II, PERCEPTORES DE EDADES 55-57 (COLUMNA 2); Y TEST DE PLACEBO A (2010-2011) Y B (2013-2014) CON EL GRUPO DE CONTROL I (COLUMNAS 3 Y 4). COEFICIENTES Y DESVIACIONES TÍPICAS (ENTRE PARÉNTESIS)

	GRUPO CONTROL 49-51	GRUPO CONTROL 55-57	PLACEBO A	PLACEBO B
Cambio legislativo (R) (β_2)	0,391*** (0,0125)	0,475*** (0,0171)	-0,179*** (0,0120)	0,175*** (0,0106)
Edad 52-54 (β_3)	-0,877*** (0,0149)	-0,0346** (0,0162)	-0,936*** (0,0127)	-0,0328*** (0,0115)
R*Edad 52-54 (β_4)	0,858*** (0,0188)	0,605*** (0,0216)	-0,00711 (0,0179)	0,0133 (0,0164)
Constante	-4,277*** (0,0333)	-4,696*** (0,0406)	-3,618*** (0,0169)	-3,824*** (0,0167)
/lntheta	-20,00 (695,0)	-18,34 (395,2)	-18,933*** (0,0199)	-18,815*** (0,011)
Theta	2,07e-09 (1,44e-06)	1,08e-08 (4,28e-06)	5,99e-09 (1,20e-10)	6,74e-09 (7,29e-11)
Log likelihood	-154.458,51	-128.760,31	-164.240,74	-150.432,89
Individuos	211.079	187.364	238.276	174.083
Episodios individuales	359.566	4.072.489	4.644.175	1.952.781

Nota: El cuadro muestra las estimaciones del modelo de duración con una variable que indica el cambio de política, otra variable que indica la pertenencia al grupo de edad de 52 a 54 años, el término de interacción de ambas variables y las variables explicativas adicionales, que incluyen atributos personales, laborales y del empleador, como sexo, nacionalidad, número de hijos, ocupación, industria, tasa de desempleo regional y 36 variables ficticias mensuales. Niveles de significación: * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Fuente: Datos del SEPE (2010-2017).

es significativa. Además, no se detectan diferencias en los coeficientes de las variables entre ambos modelos al incluir la heterogeneidad inobservada.

Todos los coeficientes tienen el signo esperado. El efecto de aumentar/reducir el derecho potencial de las prestaciones es significativo. El coeficiente de «*Edad52_54*» indica que la probabilidad de que una persona abandone el desempleo en el próximo mes es un 58 por 100 menor (coeficiente de -0,877) para las personas de 52 a 54 años que fueron elegibles para la extensión de las prestaciones por desempleo (a través del «subsido de mayores de 55 años») en relación con el grupo de control de personas de 49-51 años. Al mismo tiempo, las estimaciones puntuales para ambos grupos de edad después del cambio de política son estadísticamente significativas e indican que son un 48 por 100 mayor (coeficiente de +0,391) que antes del cambio, como se puede deducir del parámetro de la variable «*R*». Este último resultado puede deberse al hecho de que la economía estaba en recesión en 2011 y mejoró en 2013.

El coeficiente de « $R * Edad52_54$ » (b_4), que mide el impacto causal de las dobles diferencias del cambio de política, presenta el resultado esperado *a priori*, lo que sugiere que el impacto de las prestaciones ilimitadas de las PA fue relevante. El coeficiente estimado fue positivo (+0,858), por lo que la probabilidad de encontrar trabajo es 2,35 veces mayor para las personas que no tuvieron acceso a esos subsidios ilimitados de PA después de que se produjera el cambio de política y después de controlar los efectos de la mejora de la situación del mercado laboral en 2013 y las similares tasas de salida de ambos grupos de edad en promedio.

Por tanto, las reducciones en el período de derecho de las prestaciones de los trabajadores de mayor edad (que limitaban el uso de las PA para asegurar ingresos hasta la jubilación como instrumento de jubilación anticipada) están asociadas a aumentos en la salida del desempleo al empleo, *ceteris paribus*. Cuando utilizamos las estimaciones para calcular la duración promedio esperada de la percepción de las PA, los resultados son 912 días para el grupo de tratamiento y 1242 días para el grupo de control. La diferencia entre ellos sugiere que hubo una reducción de 330 días en la duración promedio debido al impacto de la reforma. La conclusión sería que la elegibilidad de prestaciones por desempleo más prolongadas a edades relativamente tempranas en el caso de los trabajadores mayores

reduce sus transiciones del desempleo al empleo y aumenta la duración del desempleo.

También se han estimado modelos separados por sexo, sectores económicos y ocupaciones (3). Los resultados (no mostrados) sugieren que la probabilidad de encontrar un trabajo asociado con el cambio de política fue de algo mayor para las mujeres, los desempleados que trabajaban anteriormente en la industria manufacturera y los trabajadores en ocupaciones no manuales, en particular aquellos en trabajos de cualificación media (ocupaciones administrativas y operadores de máquinas y ensambladores, respectivamente).

Para terminar, y como elemento adicional a los análisis anteriores, se llevan a cabo dos estrategias para probar la solidez de nuestras estimaciones de referencia. En primer lugar, mostramos que nuestros principales resultados son robustos a definiciones alternativas del grupo de control. En particular, utilizamos el grupo de control II, que está compuesto por beneficiarios de PA entre 55 y 57 años antes y después del cambio de política. En segundo lugar, se realiza una prueba de placebo. Reproducimos el análisis de una reforma placebo en enero de 2011, un año y medio antes de que se llevara a cabo la reforma real, y en enero de 2014, un año y medio después de que se llevara a cabo la reforma real. El objetivo de este ejercicio es proporcionar evidencia adicional de la falta de tendencias temporales diferenciales entre los grupos de tratamiento y de control.

Por un lado, la columna 2 del cuadro n.º 2 ofrece los resultados para la muestra completa cuando se utiliza el grupo de control II. Aunque las magnitudes de los coeficientes estimados varían, los principales hallazgos permanecen. Centrándonos en el efecto del cambio normativo, el coeficiente estimado de « $R * Edad52_54$ » es +0,605, por lo que la probabilidad de encontrar trabajo fue un 83 por 100 mayor para los desempleados que no pudieron recurrir a PA ilimitadas después del cambio de la política, después de controlar los efectos de la mejora de la situación del mercado laboral en 2013 y las similares tasas de salida de ambos grupos de edad en promedio. Por tanto, el impacto de eliminar los subsidios ilimitados para los perceptores de PA tratados fue relevante. Los resultados (no mostrados) son similares para hombres y mujeres y los tres sectores y las seis categorías ocupacionales consideradas, con las diferencias mencionadas anteriormente: el impacto de la reforma fue mayor para

los trabajadores de la industria manufacturera y los trabajadores no manuales de cualificación media.

Por otro lado, una forma alternativa de probar el supuesto de tendencias iguales sería realizar una prueba de placebo, una prueba que sirve para ganar confianza en que las entradas en el desempleo habrían continuado moviéndose en paralelo después de la fecha de la reforma si la intervención no hubiera ocurrido. Para hacer esta prueba, realizamos una estimación de dobles diferencias adicional utilizando períodos «falsos» previos y posteriores a la reforma. Se han utilizado los datos de los años 2010 y 2011, suponiendo que 2010 es el período anterior a la reforma y 2011 el período posterior a la reforma. Y se ha repetido el mismo ejercicio utilizando 2013 como período anterior a la reforma y 2014 como período posterior a la reforma. Las columnas 3 y 4 del cuadro n.º 2 muestran los resultados. No se observa ningún efecto de tratamiento significativo utilizando datos de los años anteriores a la reforma (véase el parámetro β_4 en la columna 3) ni de los años posteriores a la reforma (véase el parámetro β_4 en la columna 4).

IV. CONCLUSIONES

En las últimas décadas, la mejora de las técnicas de evaluación de impacto y la disponibilidad de bases de datos con información individual longitudinal han permitido que el volumen de estudios empíricos dirigidos a analizar el impacto de las políticas públicas del mercado de trabajo haya aumentado de manera considerable. En algunos casos, los investigadores utilizan experimentos cuasiexperimentales en los que se pueden observar cambios exógenos y, a veces, selectivos en los programas. En estos trabajos, los cambios normativos brindan a los autores la oportunidad de estudiar los efectos de las políticas del mercado de trabajo sobre las variables laborales (la duración del empleo o del desempleo, la transición hacia un determinado estado, etc.) en un entorno controlable.

En este contexto, este artículo ha revisado algunos estudios que utilizan los métodos para evaluar el impacto de políticas públicas en el caso de las políticas activas y pasivas del mercado de trabajo en España y, además, ha aplicado la técnica de dobles diferencias a un cambio de política que tuvo lugar en julio de 2012 que elevó de 52 a 55 años la edad mínima para solicitar un subsidio de desempleo ilimitado.

Por un lado, la revisión de la literatura muestra que la evaluación de las políticas del mercado de trabajo todavía no está muy extendida en el caso de España. El número de trabajos que llevan a cabo evaluaciones de impacto es bastante limitado, de modo que para algunos programas de las políticas activas o pasivas solo hay unos pocos trabajos que utilicen las técnicas de evaluación. Una de las razones puede ser la dificultad de disponer de datos adecuados (sobre todo, longitudinales), que además sean representativas, para poder llevar a cabo los análisis pertinentes. Otra es que en muchas ocasiones la simultaneidad de cambios en la normativa laboral hace casi imposible delimitar tanto los grupos de tratamiento como los grupos de control adecuados.

Por otro lado, la reforma que se analiza en el artículo, al aumentar de 52 a 55 años la edad mínima para solicitar el «subsidio para mayores de 52/55 años» en 2012, de modo que los trabajadores desempleados pudieran cobrar una PA hasta el momento en que fuesen elegibles para una pensión normal de jubilación, redujo el período de derecho ilimitado para el grupo de personas de 52 a 54 años en el momento de la pérdida del empleo, proporcionando un entorno ideal para una evaluación cuasiexperimental de la influencia de la reducción de la duración potencial de las PA en las salidas del desempleo hacia el empleo. Para ello, se ha hecho uso de los datos extraídos de los registros administrativos del SEPE español.

Nuestro trabajo empírico sigue la línea de la literatura que evalúa el impacto de la extensión / reducción de la duración potencial de las prestaciones por desempleo y sus efectos sobre las transiciones desde una PA hacia el empleo utilizando un enfoque de diferencias en diferencias (o dobles diferencias). Nuestros resultados sugieren que los cambios de este importante parámetro del período de derecho de las prestaciones provocan efectos sustanciales en el comportamiento de los trabajadores mayores. En particular, la probabilidad de salir del desempleo hacia un empleo para las personas de 52 a 54 años que dejaron de tener acceso a los subsidios ilimitados después de que se produjo el cambio de política fue 2,5 veces mayor que antes del cambio, una vez que se controla la situación económica experimentada durante el período posterior a la reforma y las diferencias de las tasas de salida medias entre los grupos de edad de 52 a 54 y de 49 a 51 años. Este hallazgo se mantiene cuando utilizamos un grupo de control alternativo

(personas de 55-57 años). Estos resultados son similares a los hallazgos de la literatura empírica que apuntan a la reducción del empleo y el aumento del desempleo y la inactividad de los trabajadores mayores afectados por programas de extensión de las prestaciones por desempleo (Kyyrä y Wilke, 2007; Lalive, 2008).

NOTAS

(*) Los autores agradecen el apoyo económico del Instituto de Estudios Fiscales para realizar esta investigación, así como al Servicio Público de Empleo Estatal (SEPE) por facilitar los datos en virtud del convenio suscrito entre la SEPE y la Universidad de Alcalá. También dan las gracias a varios participantes de seminarios y conferencias, en particular a TOMI KYRÄ, por sus valiosos comentarios y sugerencias. Evidentemente, las opiniones y análisis son responsabilidad exclusiva de los autores.

(1) La MCVL es una base de datos que contiene información longitudinal sobre una muestra representativa de la población con cualquier tipo de relación con la Seguridad Social y que ha sido ampliamente utilizada por los economistas laborales desde su puesta a disposición de los investigadores por parte de la Dirección General de Ordenación de la Seguridad Social. Puede obtenerse una visión general de la base de datos consultando DURÁN (2007), GARCÍA-PÉREZ (2008), ARRANZ y GARCÍA-SERRANO (2011) y ARRANZ, GARCÍA-SERRANO y HERNANZ (2013).

(2) El análisis realizado en esta sección se basa en un documento de trabajo más extenso (ARRANZ y GARCÍA-SERRANO, 2020).

(3) Las categorías de ocupación son las siguientes: «Trabajadores no manuales de cualificación alta» (NMCA), que incluye a directores y gerentes, profesionales científicos e intelectuales y técnicos y profesionales de nivel medio; «Trabajadores no manuales de cualificación media» (NMCM), que incluye al personal de apoyo administrativo; «Trabajadores no manuales de cualificación baja» (NMCB), que incluye a los trabajadores de servicios; «Trabajadores manuales de cualificación alta» (MAC), que incluye a agricultores y trabajadores cualificados de la construcción y la industria; «Trabajadores manuales de cualificación media» (MCM), que incluye a operadores de instalaciones y máquinas y ensambladores; y «Trabajadores manuales de cualificación baja» (MCB), que incluye a los trabajadores de ocupaciones elementales.

BIBLIOGRAFÍA

- AIReF (2020). *Incentivos a la contratación. Evaluación del gasto público 2019*. Madrid: Autoridad Independiente de Responsabilidad Fiscal.
- ARRANZ, J. M. y GARCÍA-SERRANO, C. (2011). Are the MCVL tax data useful? Ideas for mining. *Hacienda Pública Española*, 199(4), pp. 151-186.
- ARRANZ, J. M. y GARCÍA-SERRANO, C. (2020). Assistance benefits and unemployment outflows of the elderly unemployed: The impact of a law change. *Papeles de Trabajo*, 2/2020. Madrid: Instituto de Estudios Fiscales.
- ARRANZ, J. M., GARCÍA-SERRANO, C. y HERNANZ, V. (2013). How do we pursue «labormetrics»? An application using the MCVL. *Revista de Estadística Española*, 55(181), pp. 231-254.
- BANCO DE ESPAÑA (2017). *Boletín Económico*, 1/2017, pp. 19-21. Madrid: Banco de España.

- BELLONI, M., MONTICONE, C. y TRUCCHI, S. (2006). *Flexibility in retirement. A framework for the analysis and a survey of European countries*. Turín: CeRP.
- CHEN, S. y VAN DER KLAUW, W. (2008). The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, 142(2), pp. 757-784.
- CUETO, B., MAYOR, M. y SUÁREZ, P. (2017). Evaluation of the Spanish flat rate for young self-employed workers. *Small Business Economics*, 49(4), pp. 937-951.
- CUETO, B. y LÓPEZ, F. (2019). Ap. p. renticeship contract: An evaluation. *Hacienda Pública Española / Review of Public Economics*, 231(4), pp. 15-39.
- DURÁN, A. (2007). La Muestra continua de vidas laborales de la Seguridad Social. *Revista del Ministerio de Trabajo y Asuntos Sociales*, 1, pp. 231-240.
- DUVAL, R. (2003). The retirement effects of old-age pension and early retirement schemes in OECD countries. *OECD Economics Department Working Paper*, 370. Paris: OECD.
- FONT, P., IZQUIERDO, M. y PUENTE, S. (2021). The effect of subsidies to mature-age employment: A quasi-experimental analysis. *Journal of Labor Research*, 42, pp. 123-147.
- GALÁN, S. y PUENTE, S. (2015). Minimum wages: Do they really hurt young people? *The B.E. Journal of Economic Analysis and Policy*, 15(1), pp. 299-328.
- GAMBERONI, E., GRADEVA, K. y WEBER, S. (2016). Firm responses to employment subsidies: A regression discontinuity approach to the 2012 Spanish labour market reform. *Working Paper Series*, n.º 1970. European Central Bank.
- GARCÍA-PÉREZ, J. I. (2008). La Muestra continua de vidas laborales (MCVL): una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada*, 16 (E-1: n.º extraordinario), pp. 5-28.
- GARCÍA-PÉREZ, J. I. y REBOLLO, Y. (2009). The use of permanent contracts across Spanish regions: Do regional wage subsidies work? *Investigaciones Económicas*, 33(1), pp. 97-130.
- GERTLER, P. J., MARTÍNEZ, S., PREMAM, P., RAWLINGS, L. B. y VERMEERSCH, C. M. (2011). *La evaluación de impacto en la práctica*. Banco Internacional de Reconstrucción y Fomento/Banco Mundial.
- JANSEN, M. y TRONCOSO, D. (2018). *El impacto de los contratos para la formación y el aprendizaje en la inserción laboral de los jóvenes*. Madrid: FEDEA.
- JENKINS, S. P. (2005). *Survival analysis*. Manuscript. Retrieved from: <https://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf>
- KHANDER, S. K., KOOLWAL, G. B. y SAMAD, H. A. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. The World Bank.
- KUGLER, A., JIMENO, J. y HERNANZ, V. (2003). Employment consequences of restrictive permanent contracts: Evidence

from Spanish labour market reforms. *IZA Discussion Paper Series*, n.º 657. Institute for the Study of Labor (IZA).

KYYRÄ, T. y WILKE, R. A. (2007). Reduction in the long-term unemployment of the elderly: A success story from Finland. *Journal of the European Economics Association*, 5, pp. 154-182.

LACUESTA, A., IZQUIERDO, M. y PUENTE, S. (2019). Un análisis del impacto de la subida del salario mínimo interprofesional en 2017 sobre la probabilidad de perder el empleo. *Documentos Ocasionales*, n.º 1902. Madrid: Banco de España.

LALIVE, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, 142, pp. 785-806.

MÉNDEZ, I. (2013). Promoting permanent employment: Lessons from Spain. *SERIEs*, 4, pp. 175-199.

PÉREZ, C. y MORAL-ARCE, I. (2015). *Técnicas de evaluación de impacto*. Garceta.

REBOLLO, Y. y GARCÍA-PÉREZ, J. I. (2021). Evaluación de impacto de políticas activas de empleo para colectivos de difícil inserción laboral. *FEDEA Policy Papers*, 2021/11. Madrid: Fundación de Estudios de Economía Aplicada.

SCHMIEDER, J., VON WATCHER, T. y BENDER, S. (2016). The effect of unemployment benefits and nonemployment durations on wages. *American Economic Review*, 106, pp. 739-777.

Resumen

En este trabajo se analizan los primeros efectos de la implantación del ingreso mínimo vital desde su puesta en marcha en 2020. Para su estudio se utilizan los microdatos de la *Encuesta de condiciones de vida 2020* y de la encuesta *Encuesta de Integración Social de la Fundación FOESSA* de 2021, así como los registros del Ministerio de Inclusión, Seguridad Social y Migraciones. El resultado más general es una limitada eficacia, con importantes problemas de cobertura, e ineficiencia horizontal y vertical: hay un gran porcentaje de la población objetivo que no recibe la prestación, junto a otro que sin pertenecer a ella sí la recibe. La principal causa de denegación es el incumplimiento de los umbrales económicos de ingresos y patrimonio.

Palabras clave: ingreso mínimo vital, eficacia, adecuación, pobreza, exclusión.

Abstract

This paper analyses the first effects of the of the Ingreso Mínimo Vital since its launch in 2020. The study uses microdata from the ECV 2020 and the EINSFOESSA survey of 2021, as well as the administrative records of the Ministry of Inclusion, Social Security and Migration. The most general result is its limited effectiveness, with important problems of coverage, and horizontal and vertical inefficiency: there is a large percentage of the target population that does not receive the benefit, along with another that does not belong to it but does receive it. The main cause of refusal is non-compliance with economic thresholds of income and wealth.

Keywords: minimum income, take-up, adequacy, poverty, social exclusion.

JEL classification: H20, I30, I38.

EL INGRESO MÍNIMO VITAL: ADECUACIÓN Y COBERTURA

Luis AYALA CAÑÓN

UNED, Equalitas

Antonio JURADO MÁLAGA

Jesús PÉREZ MAYO

Universidad de Extremadura

I. INTRODUCCIÓN (1)

EL sistema de garantía de ingresos mínimos en España está formado por un mosaico de prestaciones muy variadas, con un entramado de distintas prestaciones dirigidas a colectivos y necesidades específicas (jubilación, maternidad, enfermedad, discapacidad, edad, emigración, actividad agraria...). Su objetivo es garantizar unos ingresos mínimos, suficientes para reducir los niveles de pobreza y de desigualdad social. La preocupación de la Unión Europea (UE) por estos dos problemas es creciente y es particularmente importante en España por su elevada magnitud en el contexto comparado.

La persistencia de altos niveles de pobreza obliga a prestar mayor atención al sistema de garantía de ingresos. En España, son dos los niveles administrativos en él implicados: la Administración Central y las administraciones autonómicas. Las rentas mínimas de inserción (RMI) fueron implantadas por las comunidades autónomas entre finales de los años ochenta y el ecuador de los noventa. La Unión Europea ya invitaba a principios de aquella década a los países miembros a desarrollar programas de rentas mínimas que garantizasen el derecho a unos recursos básicos y a servicios adecuados que ase-

gurasen una vida digna a todas las personas (Comisión Europea, 1992). Más adelante, posteriores documentos e informes promovieron cada vez con mayor énfasis el desarrollo de este tipo de programas (Comisión Europea, 2006). Estos se fueron desarrollando en varios países que hasta entonces no disponían de un programa nacional, con cierta convergencia en la definición de esquemas que trataban de unir dotación económica e inserción (Moreno, 2010). Pese a ello, al igual que en la mayoría de las políticas sociales de la UE, la heterogeneidad de estos esquemas es amplia, sobre todo en aspectos relativos a organización institucional, dotación económica y condiciones de activación (Peña-Casas y Bouget, 2014).

Los nuevos programas de renta mínima han sido, por tanto, una de las principales innovaciones de política social desde las últimas décadas del siglo XX, contribuyendo a la reestructuración de la mayoría de los Estados de bienestar europeos. Esa extensión ha sido considerablemente más limitada en el caso de España, pese a que el diagnóstico de los problemas más destacados de las prestaciones de garantía de ingresos es conocido: la complejidad de su mapa, que fue desarrollándose como respuesta a necesidades muy específicas, la persistencia

pese a ello de lagunas y obstáculos en el acceso, altos niveles de inequidad por categorías demográficas y territorios, bajos niveles de adecuación y limitada eficacia en la activación de las personas beneficiarias (Ayala *et al.*, 2016; AlReF, 2019; Aguilar y Arriba, 2020).

Tales deficiencias hacen que la respuesta al aumento de las necesidades sociales en las fases recesivas sea menos eficaz que la de otros sistemas de garantía de ingresos europeos. Los niveles de adecuación de las prestaciones durante la Gran Recesión fueron claramente inferiores a los de otros países, aunque con grandes diferencias regionales (Ayala *et al.*, 2021). En 2020 España se enfrentó a una nueva crisis sin haber completado el diseño de la última red de protección social. El drástico aumento de las necesidades de los hogares por el impacto económico de las restricciones de la actividad, las citadas lagunas del sistema, con colectivos desprotegidos, y la complejidad de los procedimientos administrativos para conseguir las prestaciones, lo hacían necesario.

El dilema era cómo combinar una respuesta inmediata al aumento de la vulnerabilidad económica de los hogares con la creación de una nueva prestación de carácter permanente, que, como en otros países europeos, contribuyera a cerrar la última red de protección económica. En junio de 2020 se puso en marcha una nueva prestación, el ingreso mínimo vital (IMV), con una novedad importante respecto al conjunto de prestaciones que articulaban previamente el sistema de garantía de ingresos, al tratarse de la primera renta mínima de ámbito nacional.

Desde su inicio, la eficacia del IMV ha sido cuestionada por su reducido número relativo de unidades beneficiarias. La premura de su puesta en marcha, sin contar con una red de gestión de las solicitudes diseñada específicamente para hogares con características distintas de los perceptores de prestaciones de la Seguridad Social, y la falta de un sistema de información que permitiera monitorizar adecuadamente la evolución del programa han hecho que el cumplimiento de algunos de esos objetivos, sobre todo la cobertura de los hogares vulnerables, haya sido hasta ahora muy limitado.

En este trabajo se analiza cómo se integra el IMV en la estructura creada por los programas de renta mínima en España pasando luego a estudiar, a partir de los datos disponibles, algunos de sus problemas de eficacia y eficiencia. El trabajo se estructura como sigue. Tras una introducción al concepto, se describe el encaje del IMV en los programas de rentas mínimas. En la tercera sección se estudian las características principales. En la cuarta se analiza la adecuación de la prestación en el entorno europeo. En la quinta sección se presentan los resultados obtenidos a partir de distintas fuentes sobre la limitada cobertura de la población vulnerable. El trabajo se cierra con una recopilación de las principales conclusiones.

II. EL ENCAJE DEL IMV EN LOS PROGRAMAS DE RENTA MÍNIMA EN ESPAÑA

El eslabón final del sistema de protección social en España es un conjunto de prestaciones que representan un último recurso para las personas con ingresos

insuficientes o sin ellos y su cuantía depende de la situación económica de la unidad beneficiaria y del límite máximo establecido. Su percepción está condicionada en muchos casos a la participación en un proceso personalizado de inserción laboral con el fin de evitar el desincentivo a la búsqueda de empleo. Su finalidad es ofrecer un nivel básico de protección mientras dure la situación de necesidad, siendo una «última red» de protección social (Arriba, 2009).

Dada la falta de iniciativa por parte del Gobierno central y la asignación de las competencias asistenciales a las comunidades autónomas en la Constitución, el primer programa de rentas mínimas español se creó en 1989 en el País Vasco (ingreso mínimo familiar), siguiendo el modelo francés de renta mínima de inserción (RMI). En los años posteriores, se fueron generalizando en el resto de las comunidades autónomas. En la segunda mitad de los años noventa, cuando ya todas las regiones contaban con un programa de renta mínima, varios trabajos demostraron que, dada la notable heterogeneidad de los programas autonómicos, hablar de política española de RMI era poco realista (Aguilar, Gaviria y Laparra, 1995; Estévez, 1998; Ayala, 2000). Posteriormente, otras críticas fueron cobrando importancia. Se apreciaban, sobre todo, problemas de articulación entre prestaciones y administraciones (Ayala *et al.*, 2016), con el surgimiento de redes paralelas más o menos complementarias y con una profunda falta de coordinación (Sanzo, 2018). Además de las RMI autonómicas, existe un conglomerado de subsidios que tratan de dar respuesta a la insuficiencia de ingresos de dis-

tintas categorías demográficas mediante niveles de protección desiguales y condiciones de acceso diferentes.

Una parte importante de la atención se ha centrado en las RMI autonómicas. Las diferencias en el grado de cobertura de la población potencialmente beneficiaria y en los niveles de protección son muy grandes y dan forma a importantes problemas de equidad interterritorial. Se ha señalado, con razón, que para una gran mayoría de la población española (el 80 por 100) no existen programas de RMI (Laparra, 2004). Antes de la crisis de 2008, los programas de RMI solamente atendían al 17 por 100 de las personas situadas bajo el umbral de pobreza severa y un tercio de los perceptores pertenecían al programa vasco (Laparra y Ayala, 2009). Con la prolongación de aquella, aumentó tanto la heterogeneidad de sus características más básicas, como las condiciones de acceso o su duración, y de la protección ofrecida, con un recorte más intenso en determinadas comunidades.

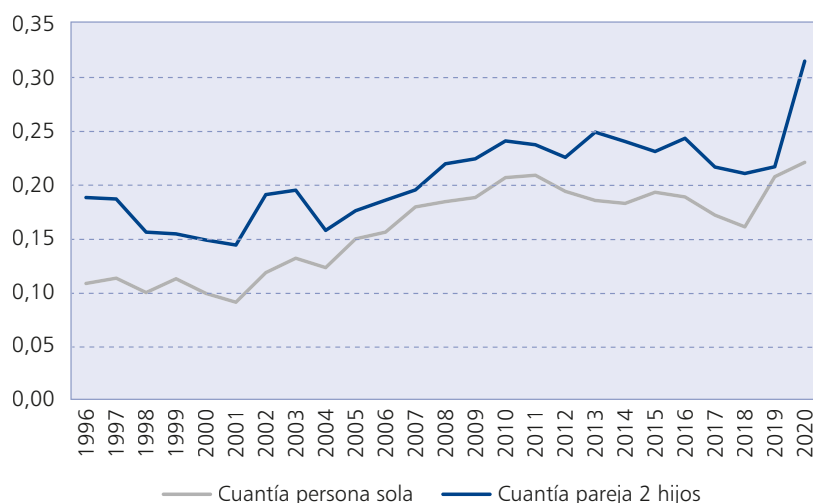
A pesar de las reformas que se han llevado a cabo en estos programas desde mediados de los años noventa, las comunidades autónomas se han encontrado con dos problemas importantes: su limitada dotación presupuestaria, con grandes dificultades para dar respuesta a las necesidades sociales en las fases recesivas, y la dificultad para disponer de una única definición del objetivo «exclusión social». Aunque la mayoría de los autores coinciden en la multidimensionalidad del problema, alcanzar un consenso respecto a los indicadores para su seguimiento, el peso de cada dimensión, los métodos de agregación y otras cuestiones meto-

dológicas, es una tarea compleja, dada la realidad autonómica actual.

Desde el año 2000 la Dirección General de Política Social, en un intento de ofrecer transparencia y facilitar una posible armonización, recopila anualmente la información básica relativa a las características de las prestaciones de las distintas RMI autonómicas. Con ello se da continuidad, en este caso oficial, a las revisiones que puntualmente y con gran limitación de datos llevaron a cabo algunos investigadores (Aguilar, Gaviria y Laparra, 1995; Ayala, 2000). El análisis de esas cifras confirma no solo la persistencia de importantes desigualdades entre las RMI autonómicas, sino su ampliación en el tiempo. Tal como muestra el cuadro n.º 1, las diferencias en los recursos invertidos por los distintos programas regionales son muy amplias.

La comunidad que asigna una mayor dotación presupuestaria a las rentas mínimas (País Vasco) gasta por habitante más de 30 veces más que la que dedica menos recursos (Murcia). Esas diferencias no se explican únicamente por la riqueza media de cada comunidad, como muestra el dato de que Madrid gaste diez veces menos, en términos relativos, que País Vasco o menos de la mitad del promedio de las RMI. Son también muy abultadas las diferencias en la incidencia demográfica de los programas, con tres regiones claramente por encima de la media (País Vasco, Navarra y Asturias) y otras muy por debajo (Madrid, Murcia y Andalucía). Esas diferencias se repiten también en las cuantías que ofrece cada programa, asociadas a niveles muy distintos de protección del riesgo de pobreza. Madrid destaca por ser la comunidad con la cuantía básica más baja, muy por debajo de la

GRÁFICO 1
DISPERSIÓN DE LAS CUANTÍAS DE LOS PROGRAMAS AUTONÓMICOS DE RMI
(Coeficiente de variación, 1996-2020)



Fuente: Elaboración propia a partir de *El Sistema Público de Servicios Sociales* (Ministerio de Derechos Sociales y Agenda 2030).

CUADRO N.º 1

DIFERENCIAS EN LOS PROGRAMAS AUTONÓMICOS DE RENTAS MÍNIMAS

	GASTO POR HABITANTE (EUROS)				BENEFICIARIOS SOBRE POBLACIÓN (%)				CUANTÍA PERSONA SOLA (EUROS)				CUANTÍA PAREJA 2 HIJOS (EUROS)			
	1996	2008	2020	1996	2008	2020	1996	2008	2020	1996	2008	2020	1996	2008	2020	2020
Andalucía	3,4	4,7	16,0	0,15	0,22	0,40	241,6	372,0	419,5	335,4	516,0	580,9	335,4	516,0	580,9	580,9
Aragón	2,4	1,8	35,8	0,09	0,07	0,64	198,3	400,0	491,0	317,3	600,0	834,7	317,3	600,0	834,7	834,7
Asturias	7,5	10,7	115,2	0,14	0,61	2,11	249,5	413,3	448,3	373,9	636,5	690,3	373,9	636,5	690,3	690,3
Baleares	1,2	3,4	19,8	0,07	0,11	0,78	245,6	383,1	461,5	349,4	613,0	731,7	349,4	613,0	731,7	731,7
Canarias	4,5	2,8	19,1	0,19	0,09	0,48	180,3	451,7	489,4	240,4	588,4	663,2	240,4	588,4	663,2	663,2
Cantabria	0,8	6,2	54,3	0,09	0,19	1,19	213,4	286,8	430,3	327,6	383,6	650,8	327,6	383,6	650,8	650,8
Castilla-La Mancha	2,5	0,7	6,2	0,12	0,02	0,13	234,1	361,8	525,0	311,4	481,2	761,3	311,4	481,2	761,3	761,3
Castilla y León	3,2	4,4	22,1	0,13	0,09	0,48	180,3	387,7	430,3	180,3	516,9	645,4	180,3	516,9	645,4	645,4
Cataluña	4,1	10,3	55,7	0,13	0,18	1,20	234,4	400,4	664,0	327,6	534,9	1.196,0	327,6	534,9	1.196,0	1.196,0
Extremadura	1,2	1,6	39,8	0,06	0,10	0,70	189,3	387,7	564,9	225,4	511,7	790,9	225,4	511,7	790,9	790,9
Galicia	3,3	6,9	19,7	0,18	0,21	0,49	213,8	387,7	403,4	292,6	542,8	597,0	292,6	542,8	597,0	597,0
Madrid	4,0	6,5	19,8	0,15	0,17	0,36	238,2	354,0	400,0	364,3	516,9	662,9	364,3	516,9	662,9	662,9
Murcia	1,2	0,6	11,4	0,05	0,03	0,39	216,4	300,0	430,3	330,6	498,0	645,4	330,6	498,0	645,4	645,4
Navarra	4,6	16,5	174,0	0,28	0,50	2,53	245,2	480,0	636,7	357,0	690,0	114,3	357,0	690,0	114,3	114,3
País Vasco	19,8	90,7	205,8	0,77	1,83	3,09	243,4	616,0	796,0	413,8	875,0	1.074,7	413,8	875,0	1.074,7	1.074,7
Rioja	1,2	1,3	42,2	0,09	0,08	0,96	235,6	361,8	430,3	311,0	361,8	672,3	311,0	361,8	672,3	672,3
C. Valenciana	1,7	0,5	50,3	0,09	0,06	0,66	249,5	375,5	630,0	252,9	425,5	864,0	252,9	425,5	864,0	864,0
TOTAL	4,2	9,3	41,7	0,17	0,24	0,78	224,1	395,3	508,9	312,4	546,6	716,2	312,4	546,6	716,2	716,2

Fuente: Elaboración propia a partir de *El Sistema Público de Servicios Sociales* (Ministerio de Derechos Sociales y Agenda 2030).

media y de las cuantías de comunidades como Cataluña, País Vasco o Comunidad Valenciana.

Estas diferencias en la protección del riesgo de pobreza han ido ampliándose en el tiempo (gráfico 1). La dispersión entre regiones, que había ido reduciéndose en los años inmediatamente posteriores a la generalización de las RMI, aumentó con el cambio de siglo y volvió a hacerlo en la primera etapa de la crisis de 2008 y en los años más recientes. La desigual respuesta a las crisis de estos programas bajo un diseño completamente descentralizado agudizó los problemas de equidad interterritorial.

Al inicio de la pandemia, el sistema de rentas mínimas era, por tanto, una realidad muy fragmentada territorialmente, con niveles de protección muy diferentes para situaciones similares de necesidad y con una cobertura, en términos de unidades perceptoras de la población sobre las elegibles, también muy desigual. A esos problemas de inequidad se le añadían muchas otras disfunciones, como la citada falta de articulación de las distintas prestaciones que componían el sistema de garantía de ingresos además de las RMI autonómicas, la persistencia de lagunas importantes en la cobertura de los diferentes grupos de población y la complejidad administrativa que dificultaba el acceso a las prestaciones de hogares económicamente vulnerables.

III. CARACTERÍSTICAS DEL IMV

La necesidad de extender el sistema de garantía de ingresos para dar cobertura al veloz aumento de la pobreza como consecuencia de la pandemia,

junto al objetivo de dar respuesta a algunas de las deficiencias estructurales que impedían una reducción más intensa de los indicadores de pobreza y desigualdad, hizo que en el año 2020 entrara en funcionamiento el IMV. El decreto ley por el que se estableció fija como su objetivo principal garantizar, a través de condiciones materiales mínimas, «la participación plena de toda la ciudadanía en la vida social y económica, rompiendo el vínculo entre ausencia estructural de recursos y falta de acceso a oportunidades en los ámbitos laboral, educativo, o social de los individuos» (Real Decreto Ley 20/2020, de 29 de mayo de 2020, y Ley 19/2021, de 20 de diciembre de 2021, por la que se establece el ingreso mínimo vital). El objetivo final es, por tanto, la inclusión social de la población excluida. Como se subraya en la normativa, esta se adaptará a las características del individuo, incentivando la formación y la búsqueda activa de empleo, resultando imprescindible la coordinación con comunidades autónomas y entidades locales y el desarrollo de itinerarios de inclusión flexibles y adaptados.

El IMV es una prestación económica mensual que pretende cubrir la diferencia entre el conjunto de ingresos del hogar y la renta garantizada determinada por la norma. Mediante la utilización de distintos baremos según tipologías de hogar, protege con mayor intensidad a los hogares monoparentales y a los menores en general y su carácter es indefinido siempre y cuando permanezcan las causas que motivaron su concesión. Como dice el decreto, es una prestación «suelo» compatible con las prestaciones autonómicas que las comunida-

des autónomas puedan conceder en concepto de rentas mínimas. El decreto ley estableció también que es el Instituto Nacional de la Seguridad Social quien se encarga del reconocimiento y control de la prestación.

IV. LA ADECUACIÓN DEL IMV

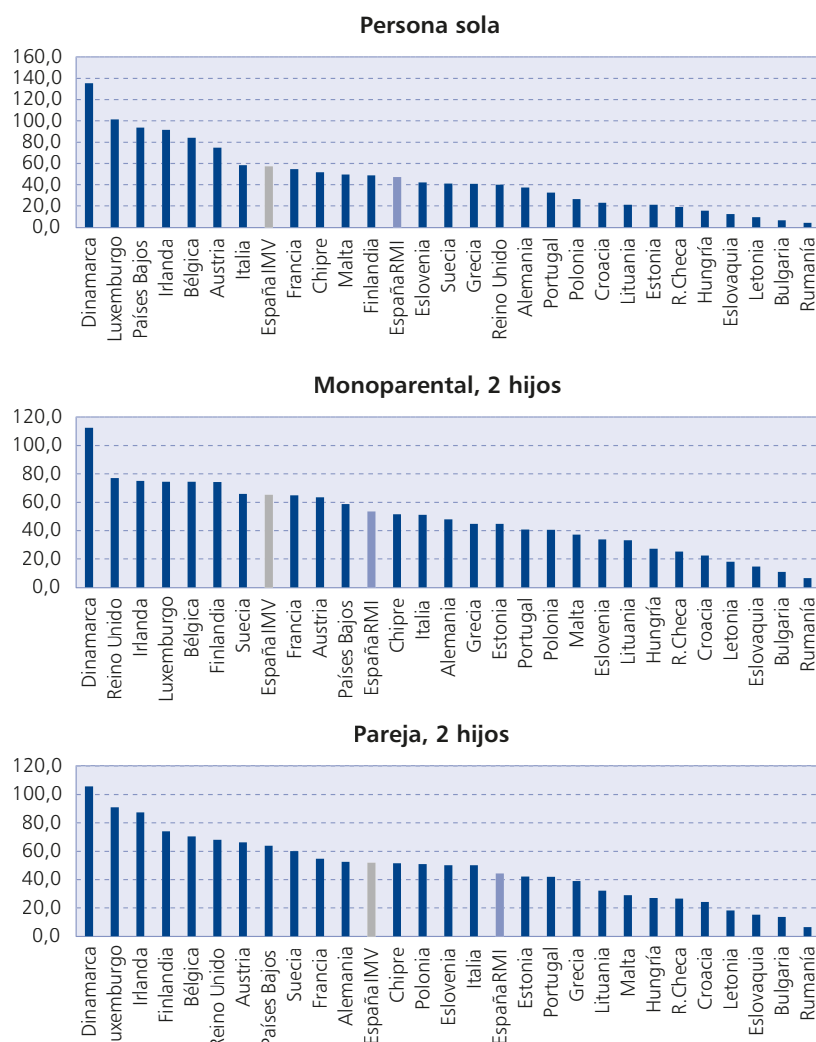
El análisis de la adecuación de la prestación es uno de los elementos habituales en la evaluación de las rentas mínimas. Como en cualquier programa de esta naturaleza, uno de los fines principales del IMV es reducir la pobreza de los hogares con menores recursos. Para algunos autores, la adecuación de una renta mínima sería la capacidad que tiene un programa de este tipo para garantizar que el crecimiento de las prestaciones no se desvíe de la tasa de crecimiento económico (Iacono, 2017). Desde otras interpretaciones, la adecuación sería la capacidad para satisfacer las necesidades básicas de las personas beneficiarias. Tal aproximación remite, sin embargo, a un concepto de pobreza absoluta que es poco frecuente en los países de renta alta. La alternativa más sencilla es utilizar umbrales relativos de pobreza monetaria. Aunque lo más común es el uso de la renta disponible del hogar como variable para construir esos umbrales, algunos autores optan por variables salariales (Immervoll, 2010; Vandenbroucke *et al.*, 2012).

Siguiendo el criterio más estándar, en este trabajo optamos por medir la adecuación como la relación que hay entre la cuantía de la renta mínima para cada tipo de hogar y el umbral de pobreza correspondiente a ese hogar (Figari *et al.*, 2013; Marchal, Marx y Van

Mechelen, 2016), definidos como un porcentaje de la renta mediana del país ajustada con algún tipo de escala de equivalencia. La principal ventaja de esta alternativa es la disponibilidad de información comparable entre países, la facilidad del cálculo y la inclusión de la pobreza relativa como uno de los indicadores propuestos por la Comisión Europea para la monitorización de la pobreza y la exclusión social. La principal objeción es la crítica habitual contra las líneas relativas de pobreza, como la sensibilidad de los resultados a la utilización de un porcentaje determinado de la mediana, la falta de consideración de las posibles diferencias interterritoriales motivadas por un diferente coste de la vida y la subjetividad en la elección del porcentaje respecto a la mediana.

Otra ventaja importante es la disponibilidad de información para calcular el indicador en los países de la Unión Europea. Por un lado, existe una fuente con información homogénea sobre la renta y las condiciones de vida de los Estados miembros (EU-SILC), que permite disponer de umbrales de pobreza para distintos tipos de hogar en todos los países. Por otro lado, también se dispone de información administrativa sobre la cuantía de las prestaciones de renta mínima en cada país (*Mutual Information System on Social Protection, MISSOC*). Debe advertirse que no se trata de comparaciones perfectamente homogéneas, ya que en varios países existen, además, otras prestaciones que complementan la renta mínima y que tienen en cuenta distintas circunstancias personales y familiares, como la edad, la enfermedad, los gastos de vivienda, las

GRÁFICO 2
ADECUACIÓN DE LAS RENTAS MÍNIMAS EN LA UNIÓN EUROPEA, 2019



Notas: 1. Los datos de RMI de España son el promedio de los programas de rentas mínimas. 2. El dato de IMV de 2020 ha sido deflactado para poder compararlo con los umbrales de pobreza de 2019, último año con información disponible (el dato de renta de EU-SILC 2020, último año disponible, recoge la renta del año anterior). 3. El dato de Reino Unido corresponde a 2018.

Fuentes: Elaboración propia a partir de EU-SILC y MISSOC.

necesidades educativas y muchas otras variables, que no siempre son fácilmente identificables en los ficheros del MISSOC.

El cálculo de los indicadores para distintos tipos de hogar y países permite situar el IMV en

el contexto de la Unión Europea (gráfico 2). Si la cuantía que se toma como referencia es la correspondiente a una persona sola puede decirse que con el nuevo programa España ha pasado a situarse en el primer tercio de los países de la UE según las tasas

de adecuación, aunque lejos de alcanzar el 60 por 100 del umbral de pobreza y de los países con prestaciones más generosas, como Dinamarca, Luxemburgo y Países Bajos. De la comparación trazada se desprende también que el IMV ofrece una protección mayor que el promedio de los esquemas autonómicos, que ubicaban a España cerca de la mediana de la distribución de los países según los niveles de adecuación.

Para los otros dos tipos de hogares considerados, familias monoparentales y parejas con dos hijos, la situación es menos favorable. Aunque en ambos casos la prestación es más generosa que la que en promedio ofrecían a sus ciudadanos más vulnerables las comunidades autónomas, los complementos correspondientes a los menores de edad son menos generosos en el IMV que en la mayoría de los países de mayor renta dentro de la UE. La nueva prestación sigue ubicando a España dentro del modelo mediterráneo de prestaciones asistenciales, con cuantías y niveles de protección inferiores a los de otros modelos, como el centroeuropeo o el anglosajón.

En síntesis, el IMV ha supuesto una mejora respecto al nivel de protección económica que ofrecían las rentas mínimas autonómicas, pero su adecuación todavía queda lejos, especialmente en los hogares de mayor dimensión, de la de prestaciones equivalentes de los países europeos de mayor renta.

V. LA COBERTURA DE LA POBLACIÓN VULNERABLE

La segunda dimensión relevante en el análisis de un programa de renta mínima es la co-

bertura del número de hogares vulnerables. Sin la disponibilidad de microdatos oficiales hasta el momento, dado que la *Encuesta de condiciones de vida (ECV)* recoge la renta del año anterior, es necesario acudir a otras fuentes. Una de las más completas, por su tamaño muestral y, sobre todo, por llegar mejor a los hogares vulnerables, es la *Encuesta de Integración Social EINSFOESSA*, elaborada desde hace años por la Fundación FOESSA para analizar la exclusión social en España. Hasta ahora se han realizado cinco encuestas (2007, 2009, 2013, 2018 y 2021). Progresivamente, se han ido ampliando los esfuerzos para obtener resultados significativos en el ámbito regional, con una muestra en la actualidad de casi 30.000 personas. Aunque el objetivo de la encuesta no es recoger información sobre la percepción del IMV, existen varias preguntas en la ola de 2021 que permiten una aproximación al número de perceptores respecto a los potenciales demandantes y, sobre todo, a las razones por las que estos no reciben la prestación.

a) *El concepto de non take-up en las rentas mínimas*

La identificación del problema de falta de acceso a la prestación (*non take-up*) es especialmente relevante en instrumentos como el ingreso mínimo vital, diseñado con el propósito de construir una última red de protección en todo el territorio nacional a partir de la cual garantizar un nivel de vida básico. La eficacia, entendida como el grado de cobertura de la población vulnerable, y la eficiencia asignativa, entendida como la asignación sin errores de la prestación, pueden verse limitadas, respectivamente, si la prestación no llega a una gran

cantidad de hogares elegibles o si llega a quienes no lo son. Esto último podría dar lugar, además, a problemas de equidad horizontal, si entre hogares con la misma necesidad algunos la reciben y otros no, y vertical, si hogares por encima del baremo fijado no la reciben, mientras que otros con rentas superiores a ese baremo sí lo hacen.

Para evaluar la cobertura es necesario partir de un concepto consensuado, con objeto de evitar un rango demasiado amplio de valores. En este trabajo seguimos la clasificación propuesta por Nelson y Nieuwenhuis (2021), que permite conocer de manera secuencial los distintos niveles de cobertura. Partiendo de una población de referencia, normalmente el conjunto de la población, se van eliminando las unidades que no pertenecen a los perfiles de riesgo hasta llegar a los beneficiarios potenciales. Al relacionar ambos colectivos, se obtiene un primer indicador, que puede acotarse más si se consideran los criterios de elegibilidad para recibir la prestación, ya que no todas las personas que podrían necesitarla pueden solicitarla. Quienes cumplen los requisitos forman la población elegible, colectivo que sirve para un tercer indicador, más restrictivo, que toma como referencia las unidades beneficiarias reales de la prestación: la falta de acceso o *non-take-up* o aquella situación en la que una persona u hogar que podría solicitar una prestación al pertenecer a la población elegible no la recibe.

Hay algunas cuestiones que deben tenerse en cuenta al aplicar esta clasificación a las situaciones reales. En primer lugar, pueden producirse errores o, incluso, fraude, por lo que algu-

nos perceptores de la prestación no deberían percibirla e incluso no ser elegibles (Sutherland et al., 2009). Además, como proponen Van Mechelen y Janssens (2017), puede darse la llamada falta de acceso terciaria cuando una persona vulnerable no tiene derecho a la prestación por las reglas de elegibilidad. Otros autores, como Goedemé y Janssens (2020), plantean una definición más amplia de la falta de acceso terciaria poniendo en relación todas las personas sin acceso a la prestación sin diferenciar entre los motivos de la falta de acceso y la población de referencia.

El motivo de la falta de acceso es un criterio seguido en la literatura para identificar sus distintos tipos. La falta de acceso primaria (Van Oorschot, 1994) se produce cuando una unidad elegible no solicita la prestación a la que tiene derecho y, por tanto, no la recibe. Se llama primaria porque es fruto de la decisión individual de los potenciales solicitantes. La falta de acceso secundaria tiene lugar cuando una unidad elegible inicia el proceso de solicitud, pero no recibe la prestación. Está asociada con los procedimientos y el funcionamiento de la Administración. Las posibles causas son varias, incluyendo la complejidad de esos procedimientos, la dificultad de demostrar la insuficiencia de recursos, los errores en la valoración y concesión, que esos procedimientos sean percibidos como degradantes o, incluso, la discrecionalidad de las personas encargadas de la valoración y concesión.

Esta clasificación según el origen de la falta de acceso fue revisada posteriormente con una nueva tipología basada en la falta de conocimiento, la decisión de no hacer la solicitud, la

no recepción de la prestación y, finalmente, la proposición de concesión negativa (Warin, 2010). Las dos primeras categorías equivalen a la falta de acceso primaria porque se producen cuando una persona potencialmente beneficiaria no sabe que puede acceder y tiene derecho a la prestación o cuando, sabiéndolo, decide no hacer efectiva la solicitud. La tercera categoría se corresponde con la falta de acceso secundaria según la clasificación antes propuesta, puesto que la persona efectivamente realiza la solicitud, aunque sin efecto positivo, y la última tiene lugar cuando la Administración no propone la concesión de la prestación a las personas potencialmente beneficiarias.

Una última propuesta para categorizar los diferentes tipos de falta de acceso toma como referencia la magnitud de esta y el tiempo durante el que no se accede a la prestación. La falta de acceso permanente se da cuando una persona no recibe la prestación durante todo el período transcurrido desde que empieza a formar parte de la población elegible hasta que deja de hacerlo. La transitoria se da cuando la prestación es concedida a una persona un tiempo después de haber realizado la solicitud, siempre que no sea retroactivo su disfrute (2). La última modalidad, la falta de acceso friccional, es muy similar a la anterior al basarse también en un acceso limitado en el tiempo, aunque centrándose en el período que transcurre entre la solicitud y la recepción de la prestación.

b) Fuentes disponibles para la evaluación del non take-up

La implementación de las distintas definiciones del problema

de la falta de acceso requiere disponer de diferentes tipos de datos. Dependiendo del concepto de falta de acceso utilizado se necesitarán diferentes tipos de información: número de beneficiarios efectivos y tamaño de la población elegible o de referencia si el objetivo es la tasa de falta de acceso, número de prestaciones sin usar o recibir como porcentaje de las necesarias para el acceso total, o tiempo total que las personas elegibles estuvieron sin recibir la prestación.

El más común de todos esos conceptos es el primero y es el que se va a utilizar en este trabajo. Las tasas se pueden calcular a partir de tres tipos de bases de datos: encuestas generales como la ECV, datos procedentes de registros administrativos o encuestas diseñadas especialmente para este objetivo.

La información en las encuestas para evaluar el acceso a un programa de rentas mínimas presenta diversos problemas. El primero es la infrarrepresentación de la población objetivo, ya que en casi todas las encuestas a hogares solo se recoge la información de la población que reside en una vivienda normalizada, por lo que las personas sin hogar o las residentes en una institución o una infravivienda no aparecen. Puede darse también cierto error de medida al estimar la población elegible en una prestación, así como la receptora, debido fundamentalmente a la infradeclaración tanto de las prestaciones recibidas como de otras variables relacionadas con la elegibilidad. Otro inconveniente es la falta de información necesaria para determinar la elegibilidad para la prestación por el desajuste en el período de referencia de los in-

gresos, a menudo el año anterior al de realización de la encuesta. Por último, hay otras desventajas, como el reducido tamaño muestral y la falta de información sobre los determinantes de la falta de respuesta. A pesar de estos problemas, los datos de encuesta son usados asiduamente para estimar la falta de acceso (Arrighi *et al.*, 2015; Bruckmeier y Wiemers, 2012; Callan y Keane, 2008; Chareyron, 2014; Chareyron y Domingues, 2018; Herber y Kalinowski, 2019; Matsaganis, Levy y Flevotomou, 2010; Riphahn, 2001; Rouwendal, 2002 o Tasseva, 2016). El principal motivo es la mayor disponibilidad de datos, junto a la facilidad para estimar el total de receptores y la población elegible en la misma base de datos.

Las encuestas diseñadas específicamente para medir la falta de acceso pueden recoger toda la información necesaria para estimar la elegibilidad y la percepción de la prestación, así como las posibles razones de aquella. Se trata, sin embargo, de datos cuya producción es cara y que requieren una cantidad de tiempo elevada para su elaboración. No están exentos, además, de los problemas de los errores de medida o de falta de cobertura de la muestra.

A diferencia de los datos de encuestas, los registros administrativos contienen mucha información precisa sobre las características de los hogares solicitantes y los beneficiarios o sus ingresos, en la que los errores de medida habituales en las encuestas son menores (3). La renta que aparece en estos registros es, además, la misma que usa la Administración Pública para determinar la elegibilidad y la concesión de la prestación. El principal in-

conveniente está relacionado con la población objetivo, ya que no se recoge información de toda la población, sino la referida a solicitantes y perceptores. La explotación de este tipo de datos, además, suele estar limitada a la Administración y el acceso de los investigadores es limitado.

La mejor opción es una combinación de encuestas y registros administrativos mediante procedimientos de *matching*. Las encuestas se usan para determinar la población elegible para la prestación y los registros administrativos para calcular las personas beneficiarias. Aunque puede haber problemas de consistencia entre las fuentes, es la mejor metodología porque aprovecha las fortalezas de ambos tipos de datos.

c) *Una estimación del non take-up en el IMV*

Debido a su reciente introducción, los datos disponibles de solicitantes, perceptores o población objetivo del IMV son todavía muy limitados. Para realizar el análisis se combinarán datos procedentes de diferentes fuentes, como la información administrativa del Ministerio de Inclusión, Seguridad Social y Migraciones, los microdatos de la ECV de 2020 y los microdatos de la *Encuesta de Integración Social* de la Fundación FOESSA (EINSFOESSA) de 2021.

La principal limitación que los datos imponen al análisis afecta a las unidades beneficiarias y las perceptoras, ya que la única información disponible es la procedente de las sucesivas notas de prensa publicadas por el ministerio. En estas notas aparecen los totales acu-

mulados de solicitudes aprobadas y beneficiarios efectivos (4) con alguna información global sobre la edad de los perceptores, el tipo de hogar y el número de beneficiarios menores de edad. Además, se proporciona información sobre el número de expedientes válidos, el de expedientes tramitados, las solicitudes resueltas (aprobadas y denegadas) y en subsanación.

Inicialmente, se seguirá como definición la falta de acceso terciaria (Goedemé y Janssens, 2020) relacionando todas las personas sin acceso a la prestación, sin diferenciar los motivos de esa falta de acceso, y la población de referencia. El acceso se medirá a partir de los titulares y beneficiarios declarados en las notas de prensa oficiales y la población de referencia a partir de diversas fuentes: las estimaciones del Ministerio de Inclusión, Seguridad Social y Migraciones publicadas al aprobar la medida, los hogares e individuos en riesgo de pobreza severa estimados con los microdatos de la ECV de 2020 y los hogares e individuos en riesgo de pobreza severa y exclusión severa según la EINSFOESSA 2021. Estas dos últimas bases de datos se han utilizado para servir de contraste a las estimaciones realizadas por la Administración. Para el propósito de este trabajo se toman como indicadores de las situaciones extremas de los hogares la exclusión social severa y la exclusión en el consumo, que son las dimensiones del indicador de exclusión social de FOESSA más relacionadas con el IMV. Los datos de la ECV, al igual que la estimación oficial del Ministerio a partir de datos fiscales, se basan en ingresos del año 2019.

CUADRO N.º 2

UNIDADES PERCEPTORAS Y BENEFICIARIOS POTENCIALES

CRITERIO	SOLICITUDES	BENEFICIARIOS	BENEFICIARIOS MENORES DE EDAD
Estimación oficial	850.000	2.300.000	690.000
Pobreza severa 30% ECV 2020	1.046.517	2.566.678	639.872
Pobreza severa 40% ECV 2020	1.761.075	4.508.261	1.118.950
Pobreza severa 30% FOESSA	1.314.711	3.646.387	1.124.910
Pobreza severa 40% FOESSA	1.824.842	5.322.778	1.683.595
Exclusión severa FOESSA	2.014.265	6.028.378	1.915.215
Exclusión consumo FOESSA	3.250.206	8.348.807	2.082.192

Fuentes: Ministerio de Inclusión, Seguridad Social y Migraciones, ECV y FOESSA.

El cuadro n.º 2, que recoge las diferentes propuestas de población objetivo para perceptores y beneficiarios, muestra cómo el criterio de pobreza severa usado por la Administración en el momento de aprobar el decreto regulador del IMV se acerca al 30 por 100 de la mediana de la renta equivalente del hogar. Puede observarse que la magnitud de la diferencia entre la estimación de la Administración y la procedente de la ECV de 2020 –con ingresos del año anterior– es reducida, aunque la cifra es casi un 20 por 100 inferior a la que resulta del umbral del 30 por 100 en la encuesta. Hay que advertir que esta última no tiene en cuenta criterios de elegibilidad exigidos por la normativa, como el patrimonio del hogar. Cabe recordar también que el umbral tradicionalmente utilizado para medir la pobreza severa no es el 30 por 100 de la mediana, sino el 40 por 100. El uso de esa línea de pobreza ofrece con los datos de la ECV de 2020 una población objetivo sensiblemente superior, que prácticamente duplica las previsiones de la Administración.

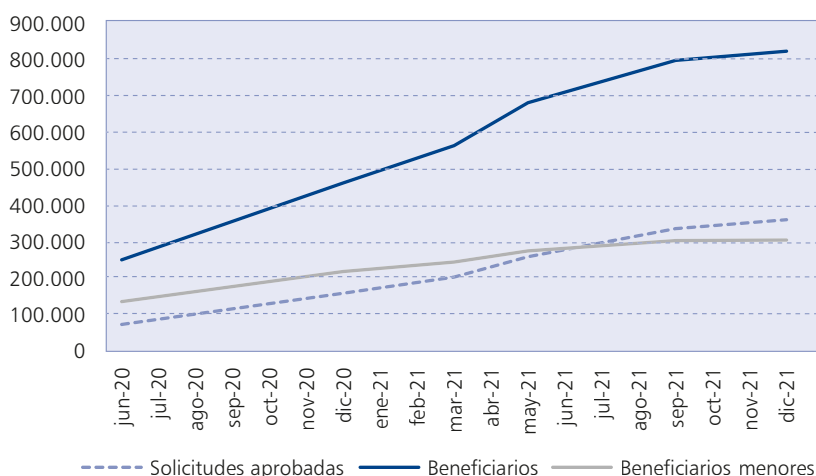
El fuerte impacto de la pandemia se constata en los indica-

dores resultantes de la encuesta EINSFOESSA, con ingresos y circunstancias del año 2021, sobre todo al comparar el indicador de pobreza severa. Los datos de exclusión severa, que incorporan dimensiones que van más allá de la renta, más que duplican las previsiones de la Administración. Este conjunto de datos muestra que las estimaciones más optimistas de la falta de acceso, en

cuanto a cifras de población más reducidas, son las oficiales.

La otra variable necesaria para determinar la falta de acceso, las personas que no acceden a la prestación, puede construirse a partir de la información sobre solicitudes aprobadas y beneficiarios procedentes de la última nota de prensa del ministerio, publicada en diciembre de 2021, un año y medio después de la implantación del IMV y justo cuando se aprobaba la ley que regula esta prestación (gráfico 3). Aunque en el primer año y medio de vigencia del IMV se produjo un paulatino crecimiento tanto de las solicitudes aprobadas como de las unidades beneficiarias, la prestación llegó únicamente a 824.441 personas desde las 362.017 solicitudes aprobadas, datos muy alejados de la previsión inicial de la Administración Pública. El ritmo de crecimiento, además, se moderó notablemente en el último trimestre de 2021. A fecha de

GRÁFICO 3
EVOLUCIÓN DE PERCEPTORES Y BENEFICIARIOS DEL IMV
(Totales acumulados)



Fuente: Ministerio de Inclusión, Seguridad Social y Migraciones.

CUADRO N.º 3

TASA DE FALTA DE ACCESO
(Porcentaje)

POBLACIÓN OBJETIVO	SOLICITUDES	BENEFICIARIOS	BENEFICIARIOS MENORES
Estimación oficial	57,4	64,2	55,7
Pobreza severa 30% ECV 2020	65,4	67,9	52,2
Pobreza severa 40% ECV 2020	79,4	81,7	72,7
Pobreza severa 30% FOESSA	72,5	77,4	72,8
Pobreza severa 40% FOESSA	80,2	84,5	81,8
Exclusión severa FOESSA	82,0	86,3	84,0
Exclusión consumo FOESSA	88,9	90,1	85,3

Fuente: Elaboración propia.

CUADRO N.º 4

GRADO DE INFORMACIÓN SOBRE EL IMV
(Porcentaje)

¿HA RECIBIDO INFORMACIÓN SOBRE EL IMV?	POBLACIÓN GENERAL	POBREZA SEVERA 30%	POBREZA SEVERA 40%	EXCLUSIÓN SEVERA
Sí, pero no tengo información suficiente para solicitarlo	3,5	11,7	14,3	13,0
Sí, y fue suficiente para solicitarlo	7,7	34,6	34,3	35,1
Sí, pero no lo necesito	16,5	4,7	5,0	8,0
No	71,8	48,5	45,8	43,4

Nota: La diferencia entre los porcentajes mostrados y el 100 por 100 se debe a la categoría no sabe/no contesta.

Fuente: EINSFOESSA 2021.

septiembre de 2021, se habían tramitado 1.244.029 solicitudes, de las que se habían denegado el 72,5 por 100.

Como puede observarse en el cuadro n.º 3, destaca la cobertura tan limitada de la prestación en sus primeros 18 meses. Incluso en las previsiones más favorables, únicamente ha llegado a un tercio de las unidades beneficiarias previstas y casi a la mitad de los hogares esperados. La utilización de otras estimaciones de la población objetivo muestra una falta de acceso generalizada, con un impacto muy reducido sobre las personas en situación de pobreza severa según los in-

gresos prepandemia (ECV 2020). Tal resultado empeora aún más cuando la identificación de la población objetivo se basa en los ingresos corrientes durante la crisis sanitaria (EINSFOESSA).

En resumen, si el objetivo principal era acabar con la pobreza severa, se puede afirmar que, por el momento, la medida ha sido poco eficaz, al haber alcanzado a una parte muy pequeña del colectivo objeto de la prestación. Esta reducida incidencia sería aún menor si el criterio para determinar la población objetivo fuera el uso de los umbrales más habituales para medir la pobreza severa o la exclusión

severa. En ese caso, la falta de acceso se elevaría al 80 por 100 de los beneficiarios potenciales.

La información disponible no permite explicar qué motiva esta falta de acceso tan elevada. Una de las razones podría ser el criterio para determinar el umbral económico, que incluye el patrimonio del hogar excluyendo la vivienda habitual y que no se puede determinar en las fuentes de datos alternativas por falta de información. En cualquier caso, el acceso es muy reducido con los datos de la Administración, que sí lo incluyó en su estimación de beneficiarios potenciales. Tampoco se dispone de información suficiente sobre el grado de conocimiento de la medida por parte de la población de referencia.

Como aproximación al análisis de los motivos de la falta de acceso, se muestra la información disponible en la última ola de la encuesta EINSFOESSA sobre la información que tienen los hogares del IMV, la dificultad del proceso de solicitud, la situación del expediente en caso de solicitud y el motivo de denegación. El conocimiento sobre la prestación, como muestra el cuadro n.º 4, parece ser insuficiente. Un porcentaje importante de los hogares declara no haber recibido información, una proporción aún más llamativa si se limita el dato a los colectivos teóricamente más interesados en el IMV, como las familias en pobreza severa con cualquiera de los dos umbrales o en exclusión severa. Como era esperable se reduce, pero manteniéndose en un valor de más del 40 por 100 de los hogares en dicha situación.

Al mismo tiempo, los datos permiten recabar otros indicios de la necesidad de mejorar la

información. Entre el 12 y el 14 por 100 de los entrevistados en la población objetivo, dependiendo del criterio, declaran haber sido informados sobre la prestación, aunque dicha información no les permitió realizar la solicitud. Destaca, asimismo, el porcentaje pequeño, pero no desdeñable, de posibles solicitantes que reportan haber recibido información a pesar de no necesitar la garantía de rentas. No hay información suficiente en la encuesta para determinar si la falta de necesidad se debe a una infradeclaración de los ingresos o a un desconocimiento real de la prestación que lleva a los hogares a pensar que no la necesitan.

Junto al desconocimiento de la prestación, otro de los motivos de la falta de acceso esgrimido en la literatura es la complejidad del proceso administrativo de solicitud. Los datos de la encuesta (cuadro n.º 5) no parecen, aparentemente, reflejar este problema, dada la reducida incidencia de los intentos de solicitud sin resultado efectivo.

Por otra parte, la especial situación motivada por la pandemia, que impulsó los procesos de digitalización de buena parte de las actividades hasta entonces presenciales, provocó que en el momento de realización de la encuesta las solicitudes por vía telemática fueran mayores que las presenciales. Destaca, de nuevo, el porcentaje de hogares que declaran no haber solicitado la prestación –más de la mitad de los que están en pobreza o exclusión severa–, de los que se desconoce el motivo de la falta de solicitud. Puede ser una decisión consciente y meditada por el hogar –falta de acceso primaria–, deberse a la falta de conocimiento que mostraba el cuadro

CUADRO N.º 5

**GRADO DE DIFICULTAD EN LA SOLICITUD DEL IMV
(Porcentaje)**

¿HA ENCONTRADO ALGUNAS DIFICULTADES EN LA SOLICITUD?	POBLACIÓN GENERAL	POBREZA SEVERA 30%	POBREZA SEVERA 40%	EXCLUSIÓN SEVERA
No se ha solicitado	89,5	54,1	53,4	55,0
Intento presencial sin tramitación	0,8	2,5	3,4	4,0
Intento telemático sin tramitación	0,7	4,6	4,5	4,3
Solicitud presencial	2,7	14,5	14,2	14,1
Solicitud telemática	4,7	22,2	22,3	21,0

Nota: La diferencia entre los porcentajes mostrados y el 100 por 100 se debe a la categoría no sabe/no contesta.

Fuente: EINSFOESSA 2021.

CUADRO N.º 6

**SITUACIÓN DE LA SOLICITUD (SOLO SOLICITANTES)
(Porcentaje)**

	POBLACIÓN GENERAL	POBREZA SEVERA 30%	POBREZA SEVERA 40%	EXCLUSIÓN SEVERA
Esperando respuesta o en trámite	31,2	36,6	35,9	36,0
Concedido, sin percepción	2,7	1,2	1,1	0,8
Percepción regular	12,8	15,1	15,2	17
Denegación	51,6	47,2	47,0	45,8

Nota: La diferencia entre los porcentajes mostrados y el 100 por 100 se debe a la categoría no sabe/no contesta.

Fuente: EINSFOESSA 2021.

n.º 4 o a la dificultad de ser atendido que lleva a no formalizar la solicitud. Finalmente, el reducido porcentaje de solicitudes permite explicar, parcialmente, las tasas de falta de acceso analizadas anteriormente.

Limitando el análisis a los solicitantes destaca, de manera importante, el gran porcentaje de solicitudes denegadas, cercano al 50 por 100, junto a la elevada frecuencia de las solicitudes sin resolución. Estos datos están condicionados por el momento de realización de la encuesta, anterior a septiembre de 2021. En esta fecha, la tasa de resolución ya se situaba en torno al 90 por 100, por lo que se puede afirmar

que la mayor parte de las solicitudes en espera reportadas en el cuadro n.º 6 fueron finalmente denegadas.

Cuando se analiza el motivo de denegación, emerge un resultado ya comentado anteriormente: se debe principalmente al incumplimiento de los umbrales económicos de ingresos y patrimonio. Es llamativo que hogares cuyos ingresos les hacen estar bajo el umbral de pobreza severa, incluso usando un criterio muy restrictivo como el 30 por 100, no cumplen los requisitos económicos. Parece necesario revisar la forma de determinar las magnitudes económicas que se deben evaluar o la declaración

CUADRO N.º 7

CAUSA PRINCIPAL DE LA DENEGACIÓN (SOLO SOLICITUDES DENEGADAS)
(Porcentaje)

	POBLACIÓN GENERAL	POBREZA SEVERA 30%	POBREZA SEVERA 40%	EXCLUSIÓN SEVERA
No cumplir con el umbral económico	61,6	43,2	47,2	56,6
No cumplir con el requisito de unidad de convivencia	13,3	13,1	15,3	13,4
Causa no indicada por la Administración	8,6	25,0	19,5	13,8
No se recuerda la causa	5,6	7,2	5,2	5,7
Otros	6,2	6,5	8,7	7,2

Nota: La diferencia entre los porcentajes mostrados y el 100 por 100 se debe a la categoría no sabe/no contesta.

Fuente: EINSFOESSA 2021.

de la situación económica en la solicitud. Llama la atención cómo las solicitudes aprobadas según la información oficial están tan alejadas de las previsiones de la propia Administración basadas en simulaciones con datos fiscales de renta y patrimonio.

Por último, los microdatos de la encuesta EINSFOESSA proporcionan un indicio del grado de ineficiencia asignativa vertical, que, como se señaló, permitía identificar la concesión de la prestación a personas u hogares

no pertenecientes a la población objetivo. Aún con las cautelas derivadas del reducido tamaño muestral, destaca el porcentaje de hogares no pertenecientes a la población objetivo que declaran recibir la prestación, coincidente con altos porcentajes de hogares que, aunque pertenecientes a la población objetivo, ven denegadas sus solicitudes por superar el umbral económico. Este tipo de errores son una importante fuente de inequidad.

VI. CONCLUSIONES

El ingreso mínimo vital (IMV) responde a una prolongada insuficiencia del Estado de bienestar en España en el ámbito de la garantía de ingresos, cubierto hasta su implantación por una amplia gama de rentas mínimas desarrolladas por las administraciones autonómicas. Estas prestaciones se caracterizan por una gran heterogeneidad, tanto en población sin recursos cubierta como en el nivel económico de protección, con problemas de articulación entre prestaciones y administraciones y una acusada falta de coordinación, que ha

supuesto niveles altos de inequidad interterritorial. En un intento de corregir esas desigualdades y de cerrar las lagunas existentes en el sistema de garantía de ingresos, se creó en 2020 el IMV como una prestación mensual que pretendía cubrir la diferencia entre los ingresos del hogar y la renta que se fijó como objetivo mínimo.

Este trabajo supone una primera valoración de algunos de sus resultados en términos de eficacia y eficiencia asignativa. Un aspecto, sin duda, positivo es la constatación de una mejora en los indicadores de adecuación respecto al nivel previamente ofrecido por los programas autonómicos. Los indicadores, sin embargo, todavía quedan lejos de los de programas similares en los países europeos de mayor renta, especialmente en el caso de los hogares de mayor dimensión.

Con la escasa información disponible, dado el breve período de desarrollo del IMV, el trabajo ha intentado también medir el grado de falta de respuesta o *non take-up* de la prestación. Mediante los microdatos de la ECV 2020 y de la encuesta EINSFOESSA 2021, junto a la información divulgada por el Ministerio de Inclusión, Seguridad Social y Migraciones, se han obtenido algunos resultados relevantes. En primer lugar, el criterio de pobreza usado por la Administración para determinar la población objetivo en el momento de aprobar el decreto regulador del IMV se asemeja al del 30 por 100 de la mediana de la renta equivalente del hogar. Si se hubiera utilizado el umbral del 40% de la mediana, más frecuente en la literatura especializada, la población objetivo casi se habría duplicado.

CUADRO N.º 8

PERCEPTORES NO PERTENECIENTES A LA POBLACIÓN OBJETIVO
(Porcentaje)

CONCESIONES	
No pobreza severa 30%	27,7
No pobreza severa 40%	38,1
No exclusión severa FOESSA	41,0

Nota: La diferencia entre los porcentajes mostrados y el 100 por 100 se debe a la categoría no sabe/no contesta.

Fuente: EINSFOESSA 2021.

Atendiendo a la falta de acceso, destaca notablemente el limitado alcance de la medida en sus primeros dieciocho meses. Solamente ha llegado a un tercio de los beneficiarios previstos y casi a la mitad de los hogares esperados, todo ello en el mejor escenario de identificación de la población objetivo. Si la meta principal es acabar con la pobreza severa, se puede afirmar que, por el momento, es una medida con una eficacia limitada. Una de las razones principales es la deficiencia en la información sobre el IMV.

De los resultados se desprende también que la principal causa de denegación es el incumplimiento de los umbrales económicos de ingresos y patrimonio. Parece necesaria la revisión de la manera de determinar los recursos económicos que se deben evaluar. Por último, se constata también la presencia de problemas de ineficiencia asignativa vertical, por el porcentaje de unidades no pertenecientes a la población objetivo que declaran recibir la prestación. Este resultado revela la existencia no solo de amplios márgenes de mejora en la asignación de las prestaciones, sino también en términos de la equidad del programa.

El IMV, en síntesis, es una buena política, necesaria para cubrir algunas de las lagunas históricas del sistema de garantía de ingresos y reducir su fragmentación territorial, pero está afectada todavía por problemas importantes de diseño y gestión que limitan su eficacia. Aunque algunas de estas deficiencias son propias de cualquier instrumento de reciente creación, hay otras que deben ser revisadas y reformuladas para conseguir una mayor reducción de los altos niveles de pobreza de la sociedad española.

NOTAS

(1) LUIS AYALA agradece la financiación recibida de la Comunidad de Madrid (H2019/HUM-5793).

(2) Este último concepto también podría extenderse a las situaciones en las que el disfrute de la prestación finaliza antes de tiempo debido a un error administrativo o a la falta de cumplimentación por parte de las unidades beneficiarias de ciertos requisitos para continuar recibiendo la.

(3) Desde el cambio metodológico de los datos de renta de la ECV, que han pasado a proceder de registros de la Seguridad Social y la Agencia Tributaria, esta desventaja de la encuesta se ha reducido.

(4) Son valores diferentes porque la solicitud se realiza por unidad de convivencia y se benefician todos los miembros de dicha unidad.

BIBLIOGRAFÍA

AGUILAR, M. y ARRIBA, A. (2020). Out of the wilderness? The coming back of the debate on minimum income in Spain and the Great Recession. *Social Policy & Administration*, 54, pp. 556-573

AGUILAR, M., GAVIRIA, M. y LAPARRA, M. (1995). *La caña y el pez. El salario social en las Comunidades Autónomas, 1989-1994*. Madrid: Fundación FOESSA.

AIReF (2019). *Los programas de rentas mínimas en España*. Madrid: Autoridad Independiente de Reforma Fiscal.

ARRIBA, A. (2009). Rentas Mínimas de Inserción de las Comunidades Autónomas: una visión conjunta de su evolución y alcance. *Gestión y Análisis de Políticas Públicas*, 2, pp. 81-100

ARRIGHI, Y., DAVIN, B., TRANNOY, A. y VENTELOU, B. (2015). The non-take up of long-term care benefit in France: A pecuniary motive? *Health Policy*, 119, pp. 1338-1348.

AYALA, L. (2000). *Las Rentas Mínimas en la reestructuración de los Estados de Bienestar*. Madrid: Consejo Económico y Social.

AYALA, L., ARRANZ, J. M., GARCÍA-SERRANO, C. y MARTÍNEZ, L. (2016). *El sistema de garantía de ingresos en España: tendencias, resultados y necesidades de reforma*. Madrid: Ministerio

de Sanidad, Servicios Sociales e Igualdad.

AYALA, L., ARRANZ, J. M., GARCÍA-SERRANO, C. y MARTÍNEZ-VIRTO, L. (2021). Towards a Fundamental Reform of Minimum Income Benefits in Spain. *International Journal of Social Welfare*, 30, pp. 152-169.

CALLAN, T. y KEANE, C. (2008). *Non-take-up of means-tested benefits: National Report for Ireland*. Dublin: Economic and Social Research Institute.

CHAREYRON, S. (2014). Le non-recours au RSA «socle seul»: l'hypothèse du patrimoine. *TEPP Research Report*, 2014-03. TEPP.

CHAREYRON, S. y DOMINGUES, P. (2018). Take-Up of Social Assistance Benefits: The Case of the French Homeless. *Review of Income and Wealth*, 64, pp. 170-191.

COMISIÓN EUROPEA (1992). Recomendación 92/441/CEE del Consejo de 24 de junio de 1992, sobre los criterios comunes relativos a los recursos y prestaciones suficientes en los sistemas de protección social, www.europa.eu/scadplus/leg/es/cha/c10609.htm

COMISIÓN EUROPEA (2006). Concerning a consultation on action at EU level to promote the active inclusion of the people furthest from the labour market, www.europa.eu.int/eur-lex/LexUriServ/LexUriServ.do?uri=COM:2006:0044:FIN:en.pdf

ESTÉVEZ, C. (1998). *Las rentas mínimas autonómicas: estudio de las normativas reguladoras y análisis jurídico de las prestaciones*. Madrid: Consejo Económico y Social.

FIGARI, F., MATSAGANIS, M. y SUTHERLAND, H. (2013). Are European Social Safety Nets Tight Enough? Coverage and adequacy of minimum income schemes in 14 EU countries. *International Journal of Social Welfare*, 22, pp. 3-14.

GOEDEMÉ, T. y JANSSENS, J. (2020). *The concept and measurement of non-take-up. An overview, with a focus on the non-take-up of social benefits*. Deliverable 9.2, Leuven, InGRID-2 project 730998 – H2020.

HERBER, S. P. y KALINOWSKI, M. (2019). Non-take-up of student financial aid—A microsimulation for Germany. <i>Education Economics</i>, 27, pp. 52-74. IACONO, R. (2017). Minimum income schemes in Europe: is there a trade-off with activation policies? <i>IZA Journal of European Labor Studies</i>, 6, pp. 1-15. IMMERVOLL, H. (2010). Minimum-Income Benefits in OECD Countries: Policy Design, Effectiveness and Challenges. <i>OECD Social, Employment and Migration Working Papers</i>, 100. LAPARRA, M. y AYALA, L. (2009). <i>El sistema de garantía de ingresos mínimos en España y la respuesta urgente que requiere la crisis social</i>. Madrid: Fundación FOESSA. MARCHAL, S., MARX, I. y VAN MECHELEN, N. (2016). Minimum income protection in the austerity tide. <i>IZA Journal of European Labor Studies</i>, 5, pp. 1-20. MATSAGANIS, M., LEVY, H. y FLEVOTOMOU, M. (2010). Non-Take Up of Social Benefits in Greece and Spain. <i>Social Policy & Administration</i> 44, pp. 827-844.	MORENO MÁRQUEZ, G. (2010). Veinte años de rentas mínimas de inserción autonómicas: el caso vasco dentro del contexto español y europeo. <i>Revista de Fomento Social</i>, 259, pp. 471-490. NELSON, K. y NIEUWENHUIS, R. (2021). Towards a new consolidated framework for analysing benefit coverage. <i>Journal of European Social Policy</i>, 31, pp. 352-362. PEÑA-CASAS, R. y BOUGET, D. (2014). Towards a European minimum income? Discussions, issues and prospects. En D. NATALI (ed.), <i>Social Developments in the European Union 2013</i>, pp. 131-159. Bruselas: ETUI. ROUWENDAL, J. (2002). Measuring and Explaining the Limited Take-up of the Housing Benefit in the Netherlands. <i>Discussion Paper Tinbergen Institut</i>, 033/3. SANZO, L. (2018). La política de garantía de ingresos en España. <i>Zerbitzuan</i>, 65, pp. 41-51. SUTHERLAND, H., DECOSTER, A., MATSAGANIS, M. y TSAKLOGLOU, P. (2009). Accurate income measurement for the	assessment of public policies. <i>Final report STREP Project n.º 028412</i>, ISER-Essex. TASSEVA, I. V. (2016). Evaluating the performance of means-tested benefits in Bulgaria. <i>Journal of Comparative Economics</i>, 44, pp. 919-935. VAN MECHELEN, N. y JANSSENS, J. (2017). Who is to blame? An overview of the factors contributing to the non-take-up of social rights. <i>Working Papers</i>, 1708. Herman Deleeck Centre for Social Policy, University of Antwerp. VAN OORSCHOT, W. (1994). <i>Take it or leave it: a study of non-take-up of social security benefits</i>. Tilburg: Tilburg University Press. VANDENBROUCKE, F., CANTILLON, B., VAN MECHELEN, N., GOEDEME, T. y VAN LANCKER, A. (2012). The EU and Minimum Income Protection: Clarifying the Policy Conundrum. <i>CSB Working Paper</i>, n.º 5/2012. Herman Deleeck Centre for Social Policy, University of Antwerp. WARIN, P. (2010). Le non-recours: définition et typologies. <i>Document de Travail</i>, 06/2010. Odenore.
--	--	--

COLABORADORES
EN ESTE NÚMERO

ARRANZ MUÑOZ, José María. Es doctor en Economía y catedrático en la Universidad de Alcalá. Sus líneas de investigación se centran en temas de microeconomía, análisis del mercado de trabajo y evaluación del impacto de políticas públicas. Ha publicado en revistas internacionales (*Labour Economics*, *British Journal of Industrial Relations*, *Transport Policy*, *Social Indicators Research*, etc.), revistas nacionales, libros y capítulos de libro. Ha participado en numerosos proyectos y contratos de instituciones nacionales (Ministerio de Empleo, Ministerio Ciencia y Tecnología, Ministerio de Economía y Competitividad) e internacionales (Comisión Europea, Eurofound). Ha sido investigador invitado en University College London, VATT y Eurofound.

ARTÉS, Joaquín. Es catedrático de Economía Aplicada en la Universidad Complutense de Madrid. Es Doctor en Economía por la Universidad de California, Irvine.

AYALA CAÑÓN, Luis. Licenciado y doctor en Ciencias Económicas y Empresariales por la Universidad Complutense de Madrid. Es catedrático de Economía en la UNED. Ha sido subdirector general del Instituto de Estudios Fiscales. Es miembro fundador del grupo EQUALITAS (Economics of Inequality and Poverty Analysis), miembro del Comité técnico de la Fundación Foessa y comisario del programa de estudios sobre desigualdad del Observatorio Social de la Fundación «la Caixa». Su labor investigadora se ha centrado en el estudio de la distribución de la renta, la política social y el mercado de trabajo.

CASADO, José María. Director de la División de Evaluación del Gasto Público de la AIReF. Doctor por la London School of Economics, máster en economía por University College London y economista titulado del Banco de España desde 2009 en la Dirección General de Economía, Estadística e Investigación. Ha sido responsable de Evaluación de Políticas Públicas de la AIReF y coordinador adjunto del programa de Revisión de Gastos Spending Review de la Institución. Profesor en el Departamento de Fundamentos del Análisis Económico del Colegio Universitario de Estudios Financieros y de la Universidad de Zaragoza, economista sénior en el Banco de Inglaterra, investigador afiliado de Fedea y economista del proyecto de investigación de la Comisión Europea AGMEMOD en el periodo 2001-2004.

CORDERO FERRERA, José Manuel. Catedrático de Economía Aplicada en la Universidad de Extremadura. Es doctor en Economía por la misma universidad. Es autor de un amplio número de artículos publicados en revistas nacionales e internacionales sobre la medición de la eficiencia de diferentes servicios públicos y dentro del ámbito de la economía de la educación. También es autor de un manual sobre esta última temática y ha sido presidente de la Asociación de Economía de la Educación (2018-2021).

GARCÍA-SERRANO, Carlos. Licenciado en Ciencias Económicas y Empresariales y doctor en Ciencias Económicas por la Universidad de Alcalá. Ocupa una plaza de catedrático de Universidad en el Departamento de Economía de la Universidad de Alcalá. Sus líneas de investigación se centran en diversos aspectos del funcionamiento del mercado de trabajo, como la movilidad laboral, el empleo temporal, el desempleo, los salarios y el impacto de políticas activas y pasivas del mercado de trabajo. Ha participado en numerosos proyectos y contratos de investigación y ha publicado artículos en revistas especializadas nacionales e internacionales, así como diversos libros y capítulos de libros.

GONZÁLEZ LÓPEZ-VALCÁRCEL, Beatriz. Catedrática de Métodos Cuantitativos en Economía y Gestión de la Universidad de Las Palmas de GC. Investigación en Economía de la Salud. Presidenta de la Asociación de Economía de la Salud (2004-2006), de la sección *Public Health Economics* de la Asociación Europea de Salud Pública, EUPHA (2011-2012) y de SESPAS (2015-2017). Actualmente, es miembro del Consejo Asesor de Sanidad (Ministerio de Sanidad), del consejo científico asesor de la AIReF, del consejo editor de *Nada Es Gratis* y de los comités científicos externos COVID de Canarias y Asturias. Patrona de la Fundación Ernest Lluch. Académica correspondiente de la Real Academia de Ciencias de Galicia.

GRAVES, Jennifer. Profesora en la Universidad Autónoma de Madrid. Doctora en Economía por la Universidad de California, Irvine. Bachelors in Economics and Statistics por la Universidad de Washington. Ha sido profesora en la Universidad de Oklahoma. Está especializada en economía de la educación y economía laboral.

JURADO MÁLAGA, Antonio. Licenciado en Ciencias Económicas por la Universidad Autónoma de Madrid, doctor en Ciencias Económicas por la Universidad de Extremadura y ejerce como profesor titular de Universidad del área de Economía Aplicada en la UEx. En los últimos años ha publicado varios artículos en revistas nacionales e internacionales de prestigio y capítulos de libros sobre pobreza, desigualdad y bienestar económico en el ámbito territorial. Ha participado como coautor en los informes Foessa sobre exclusión y desarrollo social en España 2008, 2014, 2019 y 2022, siendo obras de referencia en España sobre esta temática.

KATARYNIUK, Iván. Es responsable de la Unidad de Políticas Europeas y Globales en el Banco de España desde 2018. Anteriormente, trabajó como economista titulado en la Unidad de Economías Latinoamericanas y Emergentes en esa misma institución y como economista en la Oficina Económica del Presidente. Posee un máster en Economía y Finanzas por el CEMFI.

LÓPEZ MUÑOZ, Lucía. Es ayudante de Investigación en el Departamento de Economía Internacional y Área del Euro del Banco de España. Previamente trabajó como analista en el Instituto de Crédito Oficial. Posee un máster en Economía por la Universitat de València.

MORA-SANGUINETTI, JUAN S. Es economista titulado y miembro del Comité Nacional de Empresa en el Banco de España-Eurosistema. En el Banco de España fue también presidente de la Asociación de Economistas. Es asimismo jurista, colegiado en el Ilustre Colegio de Abogados de Madrid (ICAM) y vicepresidente de ENATIC. Con anterioridad fue economista en la OCDE (París) y profesor invitado en CUNEF, en la EAE y en la UCM. Es licenciado en Derecho y en Economía por la Universidad Carlos III de Madrid con premios en ambas licenciaturas. Posee

un máster de Investigación en Economía por el Instituto Universitario Europeo de Florencia. Se doctoró, con mención europea «cum laude» en la Universidad Carlos III de Madrid. Posee más de 70 publicaciones, incluyendo artículos en *Journal of Policy Modeling*, *Economic Modelling* o *Small Business Economics*. Ha recibido diversos premios a su carrera y a su investigación, como el Premio de la Asociación Española de Economía al mejor artículo publicado en *SERIEs* en los últimos cuatro años (2018), el Premio Secciones del ICAM (2019) o el Premio especial al Antiguo Alumno Distinguido de la Universidad Carlos III de Madrid (recibido en 2018). Es miembro del comité de expertos que redactó el Código Ético de la Universidad Carlos III de Madrid.

MORAL-ARCE, Ignacio. Pertenece al cuerpo superior de estadísticos del estado y en la actualidad trabaja como vocal asesor en el Instituto de Estudios Fiscales en las áreas de evaluación de políticas públicas. Con anterioridad ha trabajado en la universidad, el INE y la AIREF. Es doctor en Economía en el área de Econometría y tiene experiencia de consultoría nacional e internacional en temas de evaluación de políticas públicas, mercado laboral, simulación de reformas y técnicas cuantitativas, con publicaciones tanto nacionales como internacionales en estas áreas.

PEDRAJA CHAPARRO, Francisco. Doctor en Ciencias Económicas y Empresariales por la Universidad Complutense de Madrid y catedrático de Economía Aplicada (Hacienda Pública y Sistemas Fiscales). Sus principales áreas de especialización son las haciendas subcentrales y el análisis de la eficiencia en el sector público en distintos campos como educación y justicia. Ha sido miembro de diversas comisiones nacionales de reforma de la financiación de comunidades autónomas y corporaciones locales y editor asociado de Hacienda Pública Española. Es académico correspondiente de la Real Academia de Ciencias Morales y Políticas y miembro del Consejo Asesor de la AIREF.

PÉREZ, Javier J. Es director del Departamento de Economía Internacional y Área del Euro del Banco de España, donde previamente ocupó otros puestos de responsabilidad. Es miembro del Comité de Relaciones Internacionales del Eurosistema, y anteriormente lo fue del Comité de Política Monetaria y del Grupo de Trabajo sobre Finanzas Públicas. Antes de incorporarse al Banco de España, trabajó en el Banco Central Europeo, en las universidades Pablo de Olavide de Sevilla y Complutense de Madrid (en la que se doctoró), y en el ámbito de la administración autonómica en el centro de investigación central en Sevilla. Es editor asociado de la revista *Hacienda Pública Española*.

PÉREZ MAYO, Jesús. Profesor titular de Economía Aplicada en la Universidad de Extremadura. Doctor en Economía por la Universidad de Extremadura. Su trabajo investigador se centra en la privación material, pobreza multidimensional, el bienestar social y la desigualdad. Sus trabajos sobre estos temas han sido publicados en revistas y otras publicaciones de ámbito tanto nacional como internacional. Actual-

mente ocupa los cargos de director de Cáritas Regional de Extremadura y forma parte del Consejo de Cáritas Española y el patronato de la Fundación Foessa.

PINO, Eloísa del. Subdirectora de Análisis Institucional en la División de Evaluación del Gasto Público de la AIREF. Doctora en Ciencia Política (UCM) y científica titular del Instituto de Bienes y Políticas Públicas del CSIC. Ha sido investigadora visitante en la Universidad de Oxford y Ottawa, entre otras. Fue profesora de Ciencia Política y de la Administración en la URJC y en la UAM. Dirigió el Observatorio de Calidad de los Servicios en la Agencia de Evaluación de Políticas y Calidad de los Servicios (AEVAL, Ministerio de Política Territorial, 2009-2011). Ha sido directora del Gabinete de la Ministra de Sanidad, Consumo y Bienestar Social (2018-2020).

RODRÍGUEZ SÁNCHEZ, Beatriz. Es profesora de Economía Aplicada en la Universidad Complutense de Madrid. Es doctora en Economía por la Universidad de Groningen.

SANTÍN GONZÁLEZ, Daniel. Es catedrático de Economía en el Departamento de Economía Aplicada, Pública y Política de la Universidad Complutense de Madrid (UCM) e investigador del Instituto Complutense de Análisis Económico (ICAE). Es doctor en Economía por la UCM (Premio Extraordinario y Segundo Premio Nacional de Investigación Educativa *ex-aequo*). Sus líneas de investigación son la evaluación de políticas públicas, la medición de la eficiencia y la productividad, y la economía de la educación. Ha sido consultor en temas de evaluación de los Gobiernos de Perú y Paraguay y del Banco Mundial. Actualmente es editor asociado en la revista *Applied Economic Analysis*.

Sanz, Carlos. Es doctor en economía por Princeton (2016). Trabaja como economista titulado en la Dirección General de Economía y Estadística del Banco de España. Sus intereses se centran en la economía política y economía pública. Sus artículos de investigación se han publicado en revistas académicas como *Review of Economic Studies*, *Comparative Political Studies*, *Journal of Public Economics* u *Oxford Bulletin of Economics and Statistics*, entre otras.

SIMANCAS RODRÍGUEZ, Rosa. Profesora contratada doctora en el área de Fundamentos del Análisis Económico de la Universidad de Extremadura. Doctora en Economía por la misma universidad. Ha desarrollado un considerable número de investigaciones centradas en el análisis de los determinantes del rendimiento educativo o la evaluación del impacto de medidas educativas mediante técnicas de inferencia causal.

ZUCHELLI, Eugenio. Investigador Tomás y Valiente en la Universidad Autónoma de Madrid. Es doctor en Economía por la Universidad de York. Ha sido profesor en la Universidad de Lancaster. Está especializado en Economía de la Salud.

PUBLICACIONES DE FUNCAS

Últimos números publicados:

PAPELES DE ECONOMÍA ESPAÑOLA

N.º 170. Las finanzas tras la pandemia

N.º 171. Infraestructuras terrestres, transporte y movilidad de personas

PANORAMA SOCIAL

N.º 33. Los estragos sociales de la pandemia ante el horizonte post-COVID

N.º 34. El trabajo y el empleo del futuro: debates, experiencias y tendencias

CUADERNOS DE INFORMACIÓN ECONÓMICA

N.º 286. La política económica ante el brote de inflación

N.º 287. Nuevos escenarios y reformas en curso

SPANISH ECONOMIC AND FINANCIAL OUTLOOK

Vol. 11. N.º 1 Policy challenges and implications two years after the pandemic

Vol. 11. N.º 2 The war in Ukraine and implications for Spain

PAPELES DE ENERGÍA

N.º 15. Artículos seleccionados del XVI Congreso anual de la Asociación Española para la Economía Energética

N.º 16. Los riesgos del cambio climático y de la transición energética, y sus implicaciones para el sector financiero

ESTUDIOS DE LA FUNDACIÓN

N.º 98. Envejecimiento y capital social: la importancia de las redes de amigos y la participación social en el bienestar individual

LIBROS

Marketing digital y big data. Nora Lado y Daniel Peña (editores)

PRECIO DE LAS PUBLICACIONES

AÑO 2022

Publicación	Suscripción		
	Suscripción anual	Edición papel (euros)	Gastos de envío (euros)
Papeles de Economía Española	4 números	55	España 8
			Europa 40
			Resto del mundo 80
Cuadernos de Información Económica	6 números	45	España 12
			Europa 60
			Resto del mundo 120
Panorama Social	2 números	25	España 4
			Europa 20
			Resto del mundo 40
Spanish Economic and Financial Outlook	6 números	35	España 12
			Europa 60
			Resto del mundo 120
Papeles de Energía	4 números	25	España 8
			Europa 40
			Resto del mundo 80
Publicaciones no periódicas (Libros, Estudios...) disponibles solamente en formato digital gratuito.			

Los precios incluyen el IVA.

Forma de pago: domiciliación bancaria, transferencia bancaria.

Descuento editorial: 10 % a bibliotecas, librerías y agencias.

Todas nuestras publicaciones se pueden descargar, de forma gratuita, en www.funcas.es

<http://www.funcas.es/Publicaciones>

publica@funcas.es

funcas

PAPELES DE ECONOMÍA ESPAÑOLA

Últimos números publicados

- nº 150 *Competitividad en los mercados internacionales: búsqueda de ventajas*
- nº 151 *Economía y Derecho*
- nº 152 *Redes de interacción social y espacial: aplicaciones a la economía española*
- nº 153 *Economía de las ciudades*
- nº 154 *La teoría económica de las reformas fiscales: análisis y aplicaciones para España*
- nº 155 *El negocio bancario tras las expansiones cuantitativas*
- nº 156 *Los problemas del mercado de trabajo y las reformas pendientes*
- nº 157 *Análisis económico de la revolución digital*
- nº 158 *El sector exterior en la recuperación*
- nº 159 *Deporte y Economía*
- nº 160 *Medicamentos, innovación tecnológica y economía*
- nº 161 *Presente y futuro de la Seguridad Social*
- nº 162 *La gestión de la información en banca: de las finanzas del comportamiento a la inteligencia artificial*
- nº 163 *Transición hacia una economía baja en carbono en España*
- nº 164 *Crecimiento económico*
- nº 165 *Ciclos económicos*
- nº 166 *El capital humano en la economía digital*
- nº 167 *La empresa española entre la eficiencia y la desigualdad: organización, estrategias y mercados*
- nº 168 *La calidad de las instituciones y la economía española*
- nº 169 *La innovación, un desafío inaplazable*
- nº 170 *Las finanzas tras la pandemia*
- nº 171 *Infraestructuras terrestres, transporte y movilidad de personas*

