

MIDIENDO LOS IMPACTOS DE LAS POLÍTICAS ACTIVAS DE EMPLEO EN EL EMPRENDIMIENTO

Simon Charles Parker

IVEY Business School & IZA

Resumen

En este artículo se analizan cuatro políticas activas de empleo (PAE) que afectan al emprendimiento: subvenciones a la creación de empresas impulsadas por demandantes de empleo, legislación sobre discriminación positiva para grupos con problemas de inclusión en el mercado laboral, programas de educación formal en emprendimiento e intervenciones de capacitación en gestión de empresas. También se describen los distintos métodos econométricos utilizados para la evaluación del impacto de esas políticas, se revisan los trabajos más relevantes que las han aplicado, y se formulan recomendaciones sobre qué pautas debería discurrir la investigación futura para orientar el diseño de estas políticas.

Palabras clave: políticas activas de empleo, emprendimiento, evaluación.

Abstract

This article discusses four active labor market policies (ALMPs) that affect entrepreneurship: start-up grants for jobseekers, affirmative action laws, entrepreneurship education programs, and enterprise management training interventions. I discuss various econometric methods for assessing impacts of these policies; review some prominent studies which have applied them; and formulate recommendations for policy-relevant future research

Keywords: active labor market policies, entrepreneurship, evaluation.

JEL classification: J08, M13.

I. INTRODUCCIÓN

La consideración del emprendimiento como una herramienta con la que resolver diversos retos económicos y sociales actuales, es una especie de cliché para los responsables políticos. En la década de 1980, la promoción del emprendimiento era considerado la solución a los problemas de desempleo masivo y declive industrial. Hoy, con la aceleración del cambio tecnológico, es más común su consideración como un *driver* del crecimiento, de la productividad y la innovación. Sin embargo, en ambos momentos, la pregunta a la que los responsables públicos tratan de responder es la misma: ¿qué tipo de políticas públicas pueden estimular, de manera efectiva, la creación de empresas que generen valor, y qué impacto tienen en la práctica?

Este artículo pretende dar respuesta a esas preguntas. Las limitaciones de espacio nos obligan a

centrarnos solamente en un grupo de las políticas públicas que afectan a la promoción empresarial, concretamente las políticas activas de empleo (PAE)(1). En general, las PAE son intervenciones del sector público en el mercado laboral diseñadas para ayudar a encontrar un empleo a los desempleados, y a mejorar las opciones en el mercado laboral de los subempleados y grupos de trabajadores de menor empleabilidad. En concreto, en este artículo nos interesamos por un subgrupo de esas políticas que afectan tanto a emprendedores en activo como a los potenciales mediante mecanismos del mercado de trabajo. Hablaremos de cuatro grandes PAE: subvenciones para la creación de nuevas empresas por demandantes de empleo (en inglés, también denominados *welfare bridge schemes*), programas de promoción del emprendimiento de discriminación positiva, cursos de educación formal en emprendimiento y programas de adiestramiento/capacitación en gestión de empresas.

No es el propósito de este artículo hacer una revisión detallada o integral de esas políticas. El objetivo, más bien, es explicar qué hallazgos y resultados obtenidos por los especialistas del análisis económico del emprendimiento necesitan conocer los responsables de formular las políticas; qué tipos de diseño de investigación son más rigurosos y, por tanto, deben informar a los *policy-makers* a la hora de implementar esas PAE; y a qué conclusiones sobre su impacto se han llegado en función de estas estrategias de investigación, hasta la fecha. Se analizan, brevemente, los métodos econométricos utilizados para estimar los impactos de las PAE como tratamiento y se revisan los resultados generados por estas evaluaciones en la literatura existente. Una aclaración. Cuando se habla de “impacto del tratamiento” se refiere, explícitamente, a los efectos causales de una política activa determinada sobre las variables-objetivo establecidas por los responsables de su diseño. La oportunidad de este artículo corre en paralelo al creciente interés de los expertos en *management* y emprendimiento por la inferencia causal. La mayor parte de la literatura empírica previa de evaluación de las PAE se ha basado en correlaciones, por lo que el alcance de sus hallazgos para informar el diseño de las políticas es limitado. Por lo tanto, no se entra a analizar los estudios pertenecientes a esta parte de la literatura.

Desde un primer momento, es importante señalar que la medición del impacto de un tratamiento es solo uno de los vectores que los *policy-makers* utilizan para evaluar la eficacia de una determinada PAE. Una evaluación completa de la eficacia requiere un análisis de coste-beneficio integral que incorpore información adicional: costes de administración de la PAE, pérdida de eficiencia, efectos indirectos en otros agentes, etc. (véase una explicación más detallada en Parker, 2018: capítulos 18-21).

Este artículo se estructura del siguiente modo: en la sección segunda se expone qué es lo que los *policy-makers* quieren saber y cuáles son los elementos clave de los estudios econométricos de evaluación de impacto. En la tercera sección se exponen las principales estrategias de evaluación de impacto, ordenadas por la robustez de la estrategia de identificación: experimentos naturales, método de regresión discontinua, modelos de diferencias en

diferencias, *propensity score matching* y variables instrumentales. La sección cuarta contiene una breve revisión de la literatura, centrada en algunos estudios influyentes que sirven para ilustrar el uso de cada uno de esos métodos en relación con las cuatro PAE antes mencionadas. En la sección quinta se ofrecen recomendaciones prácticas sencillas para analistas y *policy-makers* para avanzar en el conocimiento de las PAE orientadas a la promoción del emprendimiento. La sección sexta presenta las conclusiones.

II ¿QUÉ QUIEREN SABER LOS RESPONSABLES DE LA FORMULACIÓN DE LAS POLÍTICAS PÚBLICAS?

¿Cuáles son las metas que pretenden alcanzar los *policy-makers* al diseñar una política de promoción empresarial? Estas metas se traducen en la(s) variable(s) dependiente(s), Y , utilizadas en los estudios econométricos de evaluación de impacto. Los tres tipos de variable independiente más comunes en estos estudios son: indicadores de la entrada o permanencia en el autoempleo/emprendimiento; medidas del desempeño de la función empresarial (por ejemplo, acerca del crecimiento, supervivencia de la empresa, creación de empleo); y motivación a la hora de tomar la decisión de convertirse en un emprendedor, la cual es la más comúnmente utilizada en las evaluaciones de los programas formativos en emprendimiento. Dado que las motivaciones no se traducen necesariamente en resultados, el tercer tipo no suele usarse, especialmente por los economistas. Algunos estudios singulares tienen en cuenta otras variables de resultado. Por ejemplo, Caliendo y Künn (2011) estiman el impacto de un plan de ayuda basado en subsidios temporales (*welfare bridge scheme*) sobre la satisfacción laboral, así como sobre la empleabilidad posterior a una experiencia de autoempleo.

Un tratamiento asociado a una política, T , es la variable independiente principal en un estudio econométrico de evaluación de impacto. También es habitual introducir distintas variables de control, o covariables, X . Si se considera que, u denota un término de error que captura las influencias no observadas en Y , un diseño de tratamiento sencillo puede

especificarse como:

$$Y_i = \beta_0 + \beta_1 T_i + X_i \gamma + u_i \quad [1]$$

El subíndice i denota una observación muestral, generalmente una persona o la realización de una oportunidad de negocio. T_i es una variable discreta, que toma el valor 1 si i es objeto de tratamiento por la política y 0 si no lo es. El parámetro central de interés es β_1 . Para evaluar el tratamiento es de interés obtener un estimador insesgado y eficiente de ese parámetro, ya que mide el impacto esperado de la política en la variable-resultado de interés. Los estudiantes universitarios aprenden que los mínimos cuadrados ordinarios (*OLS*, *ordinary least squares*) es el mejor estimador lineal insesgado (*BLUE*, *Best Linear Unbiased Estimator*), por lo que sería esperable utilizarlo (o *logit*, si Y es una variable dependiente limitada) para estimar la ecuación [1]. Sin embargo, los supuestos necesarios para que *OLS* sea *BLUE* raramente se cumplen en el caso de las PAE por los motivos que se explicarán más adelante.

Antes de entrar a considerar estimadores alternativos, debe plantearse una cuestión práctica simple, pero importante. Incluso suponiendo que un responsable de formular una política llegue a conseguir una estimación insesgada y eficiente de β_1 , cabría preguntarse: ¿Qué van a hacer con ella? El caso más sencillo se plantea cuando la estimación no es estadísticamente significativa. La evidencia, entonces, es insuficiente para justificar la puesta en marcha de la política activa, sobre todo si su implementación implica costes no recuperables. Pero ¿qué ocurre si la estimación de β_1 difiere significativamente de cero? Supongamos que es positiva y que Y es el resultado deseado (por ejemplo, un indicador de desempeño de la función empresarial). En este caso, corresponde al responsable medir los beneficios de una Y más alta si se aplica la política y compararlos con los costes soportados. Un análisis de coste-beneficio excede del alcance marcado de este artículo, por lo que el debate siguiente se limita a estudiar los problemas técnicos de la estimación de β_1 .

En este punto es necesario aclarar algunos términos. Una estimación de β_1 que mida el impacto del tratamiento para la población en general (incluyendo tanto a quienes reciben el tratamiento

como a quienes no lo reciben) se denomina “efecto promedio del tratamiento” (*ATE*, *average treatment effect*). Este valor interesará a los responsables de las políticas cuando quieran medir el impacto a escala poblacional. Otra posibilidad es que busquen medir el impacto del tratamiento *solo en las personas tratadas*. Esto es lo que se conoce como el “efecto promedio del tratamiento en los tratados” (*ATE on the treated*), o *ATET* en su forma abreviada. Su valor suele ser diferente del del *ATE*, sobre todo si quienes reciben el tratamiento difieren de quienes no lo reciben; hablaré más sobre este punto en los dos apartados siguientes. Por último, considerar también el concepto del “*ATE local*” o *LATE* (*local ATE*). Este se define como el tratamiento para el subgrupo de casos cuyas características los sitúan “cerca” del margen de tratamiento. Hablaré más de este concepto al describir los métodos de regresión discontinua en la sección tercera.

Dejando a un lado estos dos puntos, pasamos a explicar el motivo por el que la estimación de [1] con *OLS* no es aconsejable en general. Si una política (a la que nos referiremos como “tratamiento” a lo largo del artículo) tiene un efecto positivo en los individuos tratados, podría esperarse que las personas presumiblemente más beneficiadas del tratamiento encuentren un modo de ser seleccionadas. Por ejemplo, supongamos que una universidad ofrece un programa de educación empresarial (EE) y está interesada en medir cuantos nuevos proyectos de emprendimiento añadirá el programa, Y . Si codificamos por $T = 1$ los casos que siguen el programa y $T = 0$ para los que no. Podría verse que un número desproporcionado de personas con aspiraciones empresariales opten por seguir el programa. Con carácter general, esas aspiraciones no se observan y entran en el término de error en [1], es decir, u . Por lo tanto, sería previsible una correlación positiva de T y u . Así, la estimación con *OLS* de [1] estará sesgada al alza, de modo que la información a efectos del establecimiento de la política relativa a β_1 será poco fiable y potencialmente engañosa. Concretamente, el impacto del tratamiento (y, por asociación, la conveniencia de cualquier política que establece el tratamiento) estará sobrestimado. Los resultados reales en el emprendimiento acabarán siendo mucho menores de los que el *policy-maker* hubiese esperado (Fairlie, 2023).

Esto es lo que se conoce como el problema del sesgo de “autoselección”, cuya existencia es ampliamente reconocida desde hace tiempo en la literatura de políticas públicas y emprendimiento (por ejemplo, Storey, 2000). Por tanto, para separar los efectos de los tratamientos de los de selección, son necesarias otras estrategias empíricas. Hablamos de algunas de ellas en la sección siguiente, pero antes debemos analizar otras dos áreas más en las que hay que profundizar en relación con “lo que los *policy-makers* quieren saber”.

En primer lugar, los impactos de un tratamiento T sobre un indicador de resultado Y puede medirse en distintos horizontes temporales. ¿Qué plazo interesa más al *policy-maker*? ¿Los impactos a corto o a largo plazo? Los ciclos electorales podrían predisponer a los políticos a atender más a los efectos a corto plazo; sin embargo, una Administración pública políticamente neutral que trate de alcanzar la eficiencia económica podría preferir disponer de estimaciones del impacto de las políticas a largo plazo. Para ilustrar la diferencia, sigamos con el ejemplo del programa de EE. Consideremos una estimación de β_1 para medir los efectos sobre diferentes indicadores de desempeño de la función empresarial a corto plazo, por ejemplo, seis meses después de terminar el curso de educación en emprendimiento. Esta estimación diferirá en general de otra estimación de resultados a largo plazo, por ejemplo, seis años después del tratamiento. En general, es de esperar que los efectos a corto plazo sean mayores que los efectos a largo plazo (McKenzie y Woodruff, 2014; Fairlie *et al.*, 2015). Así podría ocurrir si la eficacia del tratamiento disminuya con el tiempo; por ejemplo, los estudiantes podrían olvidar lo que han aprendido sobre cómo poner en marcha un nuevo proyecto empresarial o que estos conocimientos se deprecien en el tiempo.

Distinguir entre el impacto a corto y largo plazo requiere estrategias diferentes de recolección de datos. En este sentido, hacer un seguimiento de los casos tratados y no tratados cuando ya han transcurrido varios años desde la aplicación del tratamiento, puede resultar costoso y difícil. Entre otras cosas, la pérdida de participantes en la muestra se hace más problemática cuanto más nos alejamos de la fecha del tratamiento. Por todo esto, la mayoría de los estudios tienden a proporcionar estimaciones a corto plazo en

lugar de a largo plazo. Este problema también conlleva el riesgo de sobrestimar la eficacia de la política.

En segundo lugar, ¿les interesa a los responsables de formular las políticas medir el impacto general del tratamiento en toda la población o medir el impacto en grupos específicos? Por ejemplo, los planes de discriminación positiva están diseñados para mejorar los resultados para grupos desfavorecidos (minoría) en relación con grupos favorecidos (mayoría). Un diseño de investigación riguroso exige recopilar datos sobre grupos de interés específicos y comparar las estimaciones de β_1 para esos grupos a fin de que la evaluación del impacto sea realmente informativa.

III. ¿QUÉ DISEÑOS DE INVESTIGACIÓN SON MÁS ÚTILES PARA LOS RESPONSABLES DE FORMULAR LAS POLÍTICAS?

Esta sección contiene cuatro partes, cada una dedicada a un diseño de investigación y jerarquizadas en orden decreciente según su validez interna: experimentos de campo, experimentos naturales, métodos de emparejamiento y métodos de variables instrumentales.

1. Experimentos de campo

Los experimentos de campo requieren que el analista recopile datos a partir de una muestra de interés antes de un tratamiento y que a continuación asigne el tratamiento aleatoriamente a algunos miembros (frecuentemente la mitad) de la muestra. A estos se les denomina “grupo tratado”. El resto de la muestra (“grupo de control”) no recibe el tratamiento o bien algún tipo de placebo. Una vez concluido el tratamiento, el analista registra los resultados de ambos grupos. Si el tratamiento tiene un efecto causal en los resultados, debería observarse un cambio significativo de los resultados de los miembros del grupo tratado en comparación con los miembros del grupo de control.

La principal ventaja de los experimentos de campo es que el carácter aleatorio de la asignación del tratamiento evita los problemas de autoselección; de ahí que las estimaciones de [1] aseguran que β_1 capture los efectos puros del tratamiento. Por este motivo, los experimentos de campo se consideran generalmente

como la referencia, el paradigma, de los modelos de investigación. Los *policy-makers* podrían estar interesados en obtener el valor de β_1 para grupos demográficos específicos, lo que les ayudaría a mejorar el diseño de las PAE adaptando los objetivos para estos diferentes grupos. Por ejemplo, es probable que no todos los trabajadores desempleados vayan a beneficiarse por igual de programas de formación empresarial dirigidos a facilitar su acceso al autoempleo. Incluso cabría imaginar que una determinada Administración pública realizara un estudio piloto en el que asignara cursos de formación aleatoriamente a algunos miembros de un grupo de población y no a otros, a fin de averiguar más sobre los tipos de cursos que mejor funcionan y su impacto. De forma que, si los resultados del estudio piloto apuntaran a diferencias significativas tras el tratamiento, el programa de formación podría desplegarse a gran escala. Este es un ejemplo del establecimiento de políticas basado en la evidencia y de la experimentación en el espacio de las políticas públicas. Incomprensiblemente, los políticos parecen renuentes a encargarse de estudios piloto, pese a que su implementación suele ser poco costosa y a que aportan una base más racional para la formulación de políticas que el uso de estimaciones con *OLS* sesgadas, o el “quedarnos con lo conocido” en lugar de optar por cualquier novedad.

Aunque sus ventajas son considerables, los experimentos de campo también tienen algunas limitaciones. La primera es que, en la práctica, puede resultar difícil disponer de altas tasas de participación, y que el abandono o pérdida de participantes de la muestra es un problema habitual (McKenzie y Woodruff, 2014). Esta situación reduce el poder de los contrastes econométricos y la precisión de las estimaciones de β_1 , incrementando el riesgo de errores de tipo II, es decir, no poder rechazar una hipótesis nula falsa de impacto cero. La segunda es que, en algunos contextos, como los de áreas geográficas en las que participantes tratados y no tratados interactúan entre sí, puede producirse contaminación entre los grupos; por ejemplo, cuando tratados comparten sus conocimientos con no tratados. Esta circunstancia probablemente imprimiría un sesgo a la baja a las estimaciones de β_1 . La tercera es que las intervenciones podrían generar efectos colaterales y externalidades que no solo pueden dar lugar a contaminación indirecta, sino también afectar los resultados de forma que

estos tienen un impacto más amplio y no solo limitado al impacto indicado por β_1 . Por último, podría haber problemas de validez externa, pudiendo esta estar limitada; por ejemplo, si un programa está adaptado a un contexto específico los resultados pueden no ser generalizables ni extrapolables a contextos distintos que requieran distintas intervenciones, o si puede preverse que su aplicación pueda generar distintos impactos.

2. Experimentos naturales

En algunas ocasiones poner en marcha experimentos de campo es de difícil o imposible aplicación por razones políticas, por motivos prácticos y/o por razones éticas. Por ejemplo, los responsables políticos podrían poner objeciones morales a realizar experimentos de campo de EE en los que algunos participantes (los del grupo de tratados) reciban formación mientras otros participantes (los del grupo de control) no la reciban. Y eso es así pese a los beneficios que, en términos de conocimiento para el mejor diseño de las políticas aportarían, esos experimentos. En esos casos, la segunda mejor opción es la evaluación a través de los llamados “experimentos naturales”. Con esta expresión nos referimos a acontecimientos ocurridos en el pasado que crearon, bien por accidente o por diseño, grupos separados de tratamiento y control.

Por ejemplo, pensemos en un programa de capacitación cuya elegibilidad para participar (ser objeto de tratamiento) se determina seleccionando a aquellas personas que han obtenido una puntuación por encima de un determinado umbral en una prueba creada al efecto. Los solicitantes que obtienen puntuaciones ligeramente por debajo del umbral no acceden al tratamiento, mientras que los que obtienen una puntuación algo por encima del umbral sí lo hacen, aunque cabría esperar que esos individuos, en otras condiciones, fueran similares. Los métodos denominados de “diseño de regresión discontinua” (*RDD*, *Regression Discontinuity Design*) pueden utilizarse en estos casos para analizar los efectos del programa de capacitación sobre algunos indicadores de desempeño/resultados de la función empresarial. La gran ventaja del *RDD* es que solo se requieren supuestos débiles (y comprobables) para

obtener estimaciones de impacto β_1 , probablemente libres de sesgo de selección (Lee y Lemieux, 2010). Uno de los supuestos es que los participantes no pueden manipular su puntuación para hacerla superar el umbral. Otro supuesto es la continuidad de la variable subyacente (puntuación que define la elegibilidad, en este ejemplo) que determina la pertenencia a uno de los dos grupos. Al examinar los distintos resultados en torno al umbral, el RDD estima un efecto promedio del tratamiento local (*LATE*, *local average treatment effect*).

En otros casos, los grupos de control y cuasitratamiento surgen como entidades separadas discontinuas al producirse un cambio exógeno en una política. Por ejemplo, un cambio exógeno del tipo de un programa de discriminación positiva (AA) afecta a unos grupos y no a otros, lo que puede hacer surgir una diferencia entre los grupos en términos de resultados previos y posteriores al cambio de la política. Este enfoque, que imita la estructura de un experimento de campo, es la base del diseño denominado “de diferencia en diferencias” (*DD*, *difference-in-difference*). Un supuesto clave para la identificación es que la gente no anticipa la reforma y no puede autoseleccionarse para acceder al grupo de tratados. Si bien esto podría parecer plausible en el contexto de los programas de AA, en los que la pertenencia a un grupo étnico es exógena y presumiblemente puede verificarse, más adelante veremos que no siempre es ese el caso. Contrastar la hipótesis de “tendencias paralelas” puede servir para comprobar este supuesto y aumentar la confianza en que la estimación *DD* de β_1 está capturando realmente el impacto causal de una PAE. Otro supuesto, este quizá más difícil de contrastar, es que otros eventos ocurridos en el momento de la intervención política no hayan podido afectar también al tratamiento respecto al grupo de control.

Desde 2018 hemos asistido a una explosión de la investigación sobre *DD* (véase, por ejemplo, Roth et al., 2023). Estos avances metodológicos recientes han despertado ciertos recelos sobre la fiabilidad de los estudios anteriores a 2018. Desde el punto de vista de un responsable político, los diseños de *RDD* y *DD* tienen una desventaja, y es que son históricos: solo pueden estudiarse casos específicos que ya han tenido lugar, bien por accidente o por diseño, que dieron origen a la formación, probablemente exógena, de grupos de con-

trol y cuasitratamiento. Además, a los responsables de formular las políticas podría interesarles el impacto de tratamientos de PAE novedosos, aún no llevadas a cabo. En este caso, sería altamente recomendable apostar por los experimentos de campo (quizá inicialmente en forma de estudios piloto a pequeña escala, “acceptables” políticamente).

3. Métodos de emparejamiento

A los investigadores que utilizan datos secundarios debería preocuparles que las personas que reciben el tratamiento *T* en la ecuación [1] se autoseleccionen para el tratamiento. Los datos secundarios impiden al investigador asignar personas a grupos de control y tratamiento, pero es posible imitar diseños experimentales del modo que se describe a continuación. Imaginemos un conjunto de personas, *A*, todas las cuales reciben tratamiento, y otro grupo de personas, *B*, que no lo reciben. A continuación, hacemos corresponder a cada miembro de *A* con uno o más miembros de *B* que comparten las mismas características, *C*. El emparejamiento permite que los casos en *B* simulen el resultado contrafactual de los no tratados que son equivalentes a los individuos no tratados en *A*. Existen varios métodos de emparejamiento, aunque detallarlos y describir sus ventajas e inconvenientes iría más allá del alcance de este artículo. En todos los casos, el supuesto de identificación es la independencia condicional, es decir: en función de las características observadas, los resultados del contrafactual son independientes del tratamiento.

Naturalmente, cuanto más exhaustivo sea el conjunto de características *C*, más fiable será β_1 . En la práctica, las limitaciones de los datos y la presencia de factores inobservables limitan la fiabilidad de las estimaciones del emparejamiento (Fairlie, 2023). Así, nunca puede descartarse la posibilidad, al usar métodos de emparejamiento, de que la “selección sobre variables no observables” contamine las estimaciones de nuestro parámetro de interés.

4. Variables instrumentales (VI)

En algunos casos no es posible aplicar ni diseños experimentales ni de correspondencia para evaluar

el impacto de las PAE. En estos casos, la alternativa es la de 'instrumentar' una variable potencialmente endógena T en la ecuación [1] con algunas variables "instrumentales" (VI) que afecten a la elegibilidad para el tratamiento, pero que no afecten al resultado del interés en esa ecuación, denotada por, Y . Si se hace correctamente, la estimación de IV reducirá o eliminará el sesgo en las estimaciones OLS de β_1 de las que hablamos en la sección segunda. Sin embargo, en los detalles es donde surge el problema.

Las VI deben cumplir tres condiciones: a) deben presentar una asociación fuerte con T ; b) no deben estar correlacionadas con Y (por ejemplo, incorreladas con el término de error u en [(1)]; y c) debe existir un problema de endogeneidad demostrado que deban resolver. En numerosas aplicaciones prácticas, la más difícil de cumplir es la condición b). En principio, existen "contrastes de sobreidentificación", pero implementarlas requiere aplicar varios instrumentos. Aunque el método de VI supera al de OLS , está apareciendo un consenso general, sobre todo en el análisis económico, de que el de VI es un método menos convincente para la identificación causal que los enfoques resumidos en las secciones III.1-III.3. No obstante, sigue siendo una herramienta técnica útil para estimar los efectos de las PAE.

IV. REVISIÓN DE ALGUNOS HALLAZGOS EMPÍRICOS EXTRAÍDOS DE LA LITERATURA

Para demostrar la aplicación a las PAE de los métodos que hemos visto en el apartado anterior, hacemos un breve repaso de algunos estudios recientes que los han empleado.

1. Evidencia obtenida de experimentos de campo sobre las PAE

La mayoría de los experimentos de campo en capacitación en gestión empresarial (EMT) se han llevado a cabo en países en vías de desarrollo. El motivo, en parte, es que los costes de la realización de experimentos incentivados son más bajos en países de baja renta.

A partir de estos experimentos de campo realizados en países en vías desarrollo parece emerger

un cierto consenso en cuanto a que los programas de capacitación en gestión empresarial mejoran las prácticas comerciales. Por ejemplo, a partir de evaluación aleatorizada con empresarios de microempresas llevada a cabo por un banco de República Dominicana se extrajo una especie de regla práctica según la cual un programa de formación general simplificado sobre heurística financiera básica mejoraba significativamente las prácticas financieras de las empresas, la calidad de la información financiera y los ingresos (*Drexler et al.*, 2014). Especialmente en el caso de aquellos microempresarios con escasas competencias y prácticas financieras iniciales, la formación simplificada resultó funcionar mejor que la formación "estándar" en contabilidad que suele ser más compleja (véase también *Mano et al.*, 2012). Otro experimento de campo destacable es el realizado a un grupo de iniciativas empresariales micro, pequeñas y medianas empresas en México, el cual reveló que un año de servicios de consultoría de gestión subvencionada mejoraba significativamente la productividad total de los factores y el rendimiento de los activos de las empresas (*Bruhn*, 2018). Incluso transcurridos cinco años, el número total de empleados por empresario y la factura salarial de los mismos eran significativamente más altos en las empresas tratadas que en las no tratadas. Otros experimentos de campo han resaltado la importancia de los modelos y mentores (*Brooks et al.*, 2018; *Lafortune et al.*, 2018).

Sin embargo, algunos experimentos de campo han arrojado resultados menos favorables. Por ejemplo, *Fairlie et al.* (2015) analizaron un programa estadounidense denominado *GATE* (*Growing America Through Entrepreneurship*: Hacer crecer a EE. UU. mediante el emprendimiento). *GATE* ofrecía capacitación y asesoramiento a solicitantes interesados en convertirse en autoempleados (pero no financiación). *GATE* supuso la mayor evaluación aleatorizada de capacitación y asistencia a la promoción del autoempleo realizada hasta la fecha en Estados Unidos. El coste del programa para cada participante era de 1.321 dólares americanos. Sin embargo, *Fairlie et al.* (2015) solo detectaron impactos del programa de capacitación en el corto plazo. No hubo impacto fuerte o duradero de la capacitación ni en las ventas, ganancias o en la creación de empleo, ni siquiera para aquellos que habitualmente tienen más problemas de para hacer

frente a las restricciones de liquidez, disponen de menor capital humano o sufren de discriminación en el mercado laboral (sobre todo mujeres y minorías). Los efectos se disipaban en horizontes temporales de entre 18 y 60 meses, lo que indica que GATE no consiguió incrementar las tasas de supervivencia de los nuevos autoempleados ni de los ya existentes.

Otro trabajo que también encuentra efectos muy limitados es el de Karlan y Valdivia (2011). Estos autores analizan un grupo de mujeres beneficiarias de un programa de microfinanciación en Perú (FINCA), sin encontrar impacto alguno de la capacitación en los ingresos, los beneficios o el empleo. Lo que sí observaron fue una cierta mejora en cuanto a sus conocimientos empresariales, así como un estrechamiento de sus vínculos con la entidad financiera a la que había solicitado la microfinanciación. Resumiendo las evidencias constatadas, Quinn y Woodruff (2019) concluyen que “Si bien los datos observacionales sugieren que aun siendo importante, los experimentos acerca de los impactos de los programas de capacitación en gestión empresarial han mostrado que los métodos tradicionalmente utilizados para impartir esa formación, al menos a pequeñas empresas, no son eficaces (...) la investigación apunta a que la mayoría de los programas de capacitación empresarial fracasan porque no consiguen mejorar las prácticas comerciales”.

Varios experimentos de campo han analizado los efectos combinados de los EMT y las subvenciones. Por ejemplo, Cho y Honorati (2014) realizan un metaanálisis de esos programas en países en vías de desarrollo centrándose en estudios que utilizaron diseños experimentales o cuasiexperimentales. En general, concluyen que los programas de capacitación tenían un impacto considerable y positivo en los conocimientos y prácticas comerciales entre los más jóvenes; sin embargo, el impacto era débil sobre la puesta en marcha de nuevas empresas, sobre la expansión de las existentes y sobre las rentas empresariales. También observan que el impacto de las medidas de apoyo financiero era mayor entre las mujeres, mientras que el impacto de la capacitación empresarial era mayor entre empresarios existentes. Varios estudios posteriores han identificado efectos similares de los programas de ayuda a la capacitación y financiación. Es el caso de Berge *et al.* (2015) quienes

estudian el impacto de un programa de capacitación acompañado de un paquete de subvenciones para microempresas utilizando los datos de un experimento de campo aleatorio realizado en Tanzania. Estos autores estiman que tanto las acciones formativas como las de financiación tuvieron un fuerte impacto sobre los rendimientos empresariales, las prácticas de gestión empresarial y sobre los ingresos no financieros en el caso de los emprendedores, pero no tanto en el caso de las emprendedoras.

Por último, Premand *et al.* (2016) estudiaron los efectos de la entrada aleatoria de estudiantes universitarios tunecinos en un programa de capacitación empresarial. En el estudio obtienen evidencia de efecto positivo (entre 1-4 puntos porcentuales) de las tasas de autoempleo entre los miembros del grupo de tratamiento un año después de la graduación, así como una mejora de las competencias empresariales. No obstante, el desempleo general no varió, apuntando a una intercambiabilidad del autoempleo y el empleo por cuenta ajena (compárese con Fairlie *et al.*, 2015).

2. Evidencia obtenida de los diseños de regresión discontinua (RDD) en las PAE

Ahora nos centraremos en tres estudios de capacitación para la gestión de empresas. El primero, realizado por Klinger y Schündeln (2011), empleó un RDD para analizar los efectos de un programa de formación empresarial llevado a cabo en Centroamérica. El programa estaba diseñado para mejorar las destrezas comerciales y las competencias empresariales. El RDD evita que se produzca selección endógena en el programa al utilizar una puntuación previa a efectos de admisión basada en una prueba de capacidad empresarial potencial. Los casos cuya puntuación superaba el umbral fueron admitidos al programa y tratados, mientras que los que no llegaban al umbral no fueron admitidos. Klinger y Schündeln (2011) estimaron que la participación en este programa de formación empresarial hizo aumentar significativa y sustancialmente la probabilidad de que los participantes acabaran por crear un nuevo negocio o ampliaran un negocio existente.

En el segundo, Butler *et al.* (2016) utilizaron un RDD para evaluar un programa mixto de financiación

(16.000 USD por cada nueva empresa) y ayuda técnica dirigido a fomentar la creación de nuevas empresas de base innovadoras en Argentina. La norma de selección del RDD también se aplicó a los casos cuya puntuación estaba justo por encima y justo por debajo del umbral necesario para acceder al programa. En el estudio, *Butler et al.* (2016) estimaron efectos significativos de la política en la creación, supervivencia y empleo de nuevas empresas. Dado que muchos de los que no participaron no llevaron a cabo ningún proyecto empresarial, *Butler et al.* (2016) infirieron que el programa fue un “extra”.

En el tercero, González-Uribe y Leatherbee (2018) analizaron el programa *Start-Up Chile*, un ecosistema acelerador dirigido a estimular iniciativas empresariales de rápido crecimiento en sus primeras fases. Se centraron en dos condiciones de tratamiento que suelen estar presentes en estos aceleradores empresariales: servicios “básicos” (es decir, financiación y acceso a espacio de trabajo compartido) y formación. El método de corte de su RDD, necesario dado que el programa de *Start-Up Chile* solo aceptaba un número limitado de participantes, se basó también en una puntuación asignada a los solicitantes. González-Uribe y Leatherbee (2018) estimaron que la formación combinada con los servicios básicos hizo aumentar de forma significativa el rendimiento de nuevas iniciativas empresariales. En cambio, los servicios básicos no parecieron afectar al rendimiento de nuevas iniciativas empresariales por sí mismos.

En resumen, la evidencia obtenida a partir de los RDD en los programas de formación en gestión de empresas tiende a respaldar la eficacia esas PAE con formación en gestión empresarial.

3. Evidencia obtenida de los diseños de diferencia en diferencias (DD) en PAE de discriminación positiva

Los diseños *DD* han sido utilizados para analizar diversas PAE orientadas a la promoción. del emprendimiento, A continuación, de forma resumida, presentamos aquí dos estudios de evaluación de impacto de dos programas de promoción del autoempleo de discriminación positiva (AA). Podría argumentarse que los programas de AA pueden

considerarse PAE en cuanto a que utilizan normas de contratación pública para sesgar la demanda de productos y servicios de grupos marginales o excluidos, haciendo aumentar el atractivo del autoempleo como opción de ocupación para esos grupos.

Primero, Fairlie y Marion (2012) estimaron un modelo de triple *DDD* para comparar: i) las tasas de empleo por cuenta propia de minorías étnicas y mujeres en relación con hombres de raza blanca; ii) estados de EE. UU. con tratamiento (es decir, AA) frente a no tratamiento (no AA; y iii) antes y después de eliminar los programas de AA. Los estados con tratamiento eran California y Washington. En ambos estados se observó un efecto significativo y positivo del tratamiento en minorías étnicas y mujeres, aunque los resultados dependían en cierta medida de la elección de estados de grupo de control. El impacto positivo puede resultar sorprendente; sin embargo, podría ocurrir que las oportunidades de empleo para minorías y mujeres empeoraran al finalizar la AA, empujándolos al autoempleo.

Más tarde, Chatterji et al. (2014) aprovecharon la introducción escalonada de la provisión de fondos de los programas en ciudades de Estados Unidos en la década de 1980 como estrategia de identificación. Sus estimaciones de triple *DDD* sugerían que el porcentaje de propietarios de empresas en varones de raza negra aumentaron significativamente después de introducirse el programa, y que la brecha con respecto al mismo porcentaje entre hombres de raza blanca descendía 3 puntos porcentuales, especialmente en sectores muy afectados por los programas. Interpretan estos autores que parece que hubo una reasignación del empleo por cuenta propia entre los americanos de raza blanca a los de raza negra, ya que el porcentaje total de propietarios de empresas experimentó pocos cambios en general. Al parecer, el programa benefició sobre todo a varones negros con un mayor nivel de formación.

4. Evidencia obtenida a partir de modelos de emparejamiento (PSM o propensity score matching) y variables instrumentales (VI) en las PAE

A continuación, hablaremos de dos grandes PAE: las subvenciones para la creación de nuevas empresas

por parte de demandantes de empleo y los programas de educación formal en emprendimiento (EE).

Subvenciones para la creación de nuevas empresas por demandantes de empleo

Existe mucha literatura sobre subvenciones para la creación de empresas diseñadas para que los demandantes de empleo accedan al trabajo por cuenta propia. En general, esos programas permiten a personas desempleadas convertir pagos futuros de seguros por desempleo en una subvención y/o prestación para la financiación de una nueva empresa (capitalización del subsidio). Un reciente trabajo de la OCDE (OECD, 2023: capítulo 9) resume de manera general y comparada algunos de estos programas en varios países. Más abajo hablaremos de estudios que han utilizado métodos de *propensity score matching* para estimar los efectos de esas PAE en la transición del desempleo al autoempleo, además de en otros resultados económicos.

Sin duda, la mayor parte de la atención en esta área se ha centrado en el programa de subsidios a la entrada al autoempleo alemán (SUS). En 2011, un total de 134.000 alemanes desempleados recibieron ayuda financiera con el SUS para convertirse en empresarios. Parece que los participantes del programa se autoseleccionaron entre las personas desempleadas más aptas en términos de capital humano. El resultado fue que las tasas de supervivencia de las empresas entre los destinatarios del programa resultaron relativamente altas, con unos costes en términos de “pérdida de peso muerto” moderados (Caliendo *et al.*, 2015). Numerosos estudios que utilizan métodos de PSM también revelan efectos positivos del SUS en términos de empleabilidad e ingresos posteriores. En esta dirección apuntan los trabajos de Hinz y Junbauer-Gans (1999), Pfeiffer y Reize (2000), Baumgartner y Caliendo (2008), Caliendo (2009), Caliendo y Kritikos (2010), Caliendo y Künn (2011, 2015), Caliendo *et al.* (2015, 2016), Wolff *et al.* (2015) y Caliendo y Tübbicke (2020), entre otros.

Los métodos de PSM se han empleado también para estimar el impacto en otras variables de resultados, como la creación de empleo por parte de los empresarios tratados (Caliendo y Kritikos, 2010) y la

incorporación de mujeres desempleadas al mercado de trabajo (Caliendo y Künn, 2015). Asimismo, se han estimado efectos positivos sobre el empleo en las PAE de seguimiento denominadas Einstiegs geld (Wolff *et al.*, 2015) y Gründungszuschuss (Caliendo *et al.*, 2016). Una salvedad que debe mencionarse es que el impacto de esos programas parece variar en función de las condiciones económicas locales (Caliendo y Künn, 2017). Sin embargo, en países distintos de Alemania, la mayor parte de la evidencia sugiere que el impacto de las PAE dirigidas a desempleados con subsidios para la creación de empresas ha sido limitado para los participantes en el programa (Parker, 2018). Comentaremos más sobre este punto en la sección siguiente.

Por último, a efectos de exhaustividad, cabe señalar otras dos PAE también puestas en marcha en Alemania. Una de ellas consiste en la concesión de ayudas y subvenciones para nuevas empresas en sectores intensivos en alta tecnología y conocimiento. Hottenrott y Richstein (2020) utilizaron métodos de ajuste de PSM y estimaron que tanto las ayudas como los préstamos subvencionados facilitan la inversión en activos tangibles, el empleo y el crecimiento de los ingresos de los destinatarios. La otra política activa es la ayuda de emergencia para autónomos implementada en Alemania durante la pandemia de COVID-19. El Gobierno alemán anunció un programa de ayuda de emergencia de 50.000 millones de euros en marzo de 2020 consistente en pagos únicos de hasta 15.000 euros a trabajadores autónomos que experimentaran descensos sustanciales de sus ingresos. Butler *et al.* (2022) y Bertschek *et al.* (2024) utilizaron un PSM para estimar los efectos de ese programa sobre la confianza de los empresarios en sobrevivir a la crisis.

5. Programas de EE: PSM y IV

A diferencia de muchos otros estudios sobre EE, Elert *et al.* (2015) realizaron una evaluación de impacto a largo plazo (hasta dieciseis años después del tratamiento) de una iniciativa de EE muy conocida, el Programa Junior Achievement Company (JACP). Aplicando métodos de PSM, Elert *et al.* (2015) estimaron que la participación en el JACP

sueco hizo aumentar, a largo plazo, la probabilidad de emprender y los ingresos de los participantes, pero no tuvo efectos significativos en la supervivencia de las empresas.

Oosterbeck *et al.* (2010) estimaron efectos distintos aplicando métodos de VI. Estos autores también analizaron un programa de *Junior Achievement*, concretamente el programa de mini-empresas de estudiantes *Junior Achievement Young Enterprise* (JAYE). JAYE es un programa destacado de EE aplicado en centros de enseñanza secundaria y universidades de Estados Unidos y Europa. Es eminentemente práctico y hace hincapié en la creación y operación (por grupos de alumnos) de iniciativas empresariales de pequeño tamaño y limitadas en el tiempo. Si bien no hicieron una asignación estrictamente aleatoria de JAYE a los sujetos, Oosterbeck *et al.* (2010) aprovecharon el hecho de que el programa se implementó en una de las sedes de un colegio en Holanda, pero no en otra, instrumentando la elección de sede (y, por ende, el tratamiento del programa) por la distancia relativa de las sedes a los lugares de residencia de los padres. A continuación realizaron una regresión DD en la que las diferencias eran “tratamiento en relación con el grupo de control” y “antes en relación con después” de la implementación del programa. Sus conclusiones fueron sorprendentes: se estimó que el tratamiento de JAYE no tuvo efectos significativos en cuanto a cambios de características o competencias empresariales de los participantes. Aún peor: la participación en JAYE tuvo un efecto negativo significativo en las intenciones declaradas por los participantes de convertirse en empresarios en el futuro (véase también Von Graevenitz *et al.*, 2010). Esto sugiere que ese programa de EE podría haber disuadido a los estudiantes de plantearse la creación de una empresa como una elección profesional adecuada para ellos.

V. RESUMEN DE RECOMENDACIONES

Concluye este breve repaso a las diversas estrategias empíricas para estimar el impacto de las PAE sobre el emprendimiento con varias recomendaciones útiles para la agenda de investigación futura. El objetivo de las recomendaciones es orientar a los

académicos y a los responsables de las políticas públicas que estén interesados en acometer investigación de alta calidad sobre los efectos de las PAE.

Recomendación 1: Financiar y realizar más experimentos de campo

Si realmente queremos averiguar los efectos de las PAE, tenemos que realizar diseños de investigación cuyos niveles de validez interna sean altos; idealmente, pruebas de control aleatorias. Hemos visto cómo, aunque este enfoque se ha utilizado en algunos programas de formación de empresas y educación empresarial, esos programas se han llevado a cabo en países en desarrollo y para una tipología reducida de PAE. Así, por ejemplo, los programas diseñados para que personas desempleadas transiten hacia iniciativas empresariales no han sido estudiados exhaustivamente mediante experimentos de campo. En su lugar, los responsables de establecer las políticas han tenido que basarse en estudios que, en el mejor de los casos, han empleado métodos de emparejamiento. Aunque este método es mejor que ninguno, lo cierto es que formula hipótesis fuertes sobre la comparabilidad de grupos de control y cuas-tratamiento basándose en características observables. Dicho brevemente: no es ideal como enfoque metodológico.

Un perspicaz estudio de Fairlie (2023) ilustra este punto de un modo crítico. A diferencia de la mayor parte de los estudios y que utilizan uno o, a lo sumo, dos métodos para estimar los efectos del tratamiento, Fairlie (2023) comparó métodos experimentales y no experimentales empleando un mismo conjunto de datos. En el apartado 1 de la sección cuarta se menciona el programa GATE: analizándolo como experimento de campo, Fairlie (2023) concluyó que los efectos del tratamiento de formación empresarial en el porcentaje de propietarios de empresas, las ventas y la contratación fueron insignificantes. En cambio, las estimaciones no experimentales que aplican un conjunto de controles especialmente amplio, incluidas características normalmente no observables tales como experiencia empresarial familiar previa, restricciones crediticias y rasgos de personalidad, arrojaron efectos considerables, positivos y estadísticamente significativos. Resulta intere-

sante que Fairlie investigara el rendimiento de los métodos de PSM para concluir que también generaba estimaciones sesgadas y considerablemente infladas de impactos, positivos y estadísticamente significativos. De lo que se deduce, en pocas palabras, que los métodos importan. Si los responsables políticos quieren tomar decisiones basadas en evidencias, deben encargar estudios que utilicen diseños de investigación fiables. Idealmente, experimentos de campo.

En este punto, un *policy-maker* podría poner objeciones basadas en lo utópico de esa idea: los experimentos de campo son demasiado caros y plantean complicaciones relacionadas con el aspecto ético de experimentar con sujetos humanos. Discrepamos de esa opinión. En primer lugar, los estudios piloto pueden implementarse a bajo coste y la asignación aleatoria del tratamiento suele ser directa. En segundo lugar, no debemos ser tan escrupulosos a la hora de ofrecer oportunidades de mejoras de Pareto a unos miembros de la población sin perjudicar a los miembros del grupo de control, sobre todo cuando los políticos suelen mostrar pocos escrúpulos cuando se trata de experimentación con medicamentos potencialmente peligrosos en sujetos humanos por parte de empresas farmacéuticas. Para hacerlo realidad, haría falta una mentalidad atrevida, imaginativa y nueva en los círculos políticos. Y, desde luego, tampoco creo que a los responsables políticos que lean este artículo deba preocuparles que el número de investigadores dispuestos a colaborar con un gobierno que encargue estudios sobre las PAE basados en un diseño de investigación de experimentos de campo, sea escaso.

Recomendación 2: Prestar más atención a los efectos heterogéneos del tratamiento

Resulta tentador esperar respuestas simples para preguntas complicadas. No obstante, la experiencia nos enseña que eso suele ser poco realista o menos útil de lo que parece a simple vista. Una “respuesta simple” en el contexto de las PAE sería algo así: “La exposición a la política activa examinada afecta los resultados de la empresa Y en un Z por ciento”. Aunque esa respuesta se considere un ATE, ATET o LATE (para emplear la terminología

definida en la sección segunda), la idea de que una política aplicable a una población enormemente diversa pueda basarse en un solo número resulta difícilmente creíble. Aun así, esa idea persiste en muchos de los estudios publicados. La verdad es que las muestras de pequeño tamaño suelen impedir al investigador modelar la heterogeneidad y solo le permiten informar de una estimación de impacto media y única. Pero un número único invariablemente oculta impactos heterogéneos en distintos grupos de personas afectados de formas diferentes por la política.

Idealmente, el investigador dispondría de datos suficientes para estimar los impactos del tratamiento de una política activa en varios grupos de interés distintos. Si así fuera, podría revelarse, por ejemplo, que una determinada política es más eficaz cuando se dirige a un grupo que cuando se dirige a otro. Por ejemplo, Rosas *et al.* (2022) llevaron a cabo un ensayo de control aleatorio en el que las personas tratadas recibían tanto un flujo regular de pequeñas inyecciones de efectivo como formación en competencias técnicas, en competencias empresariales o en una combinación de ambas. Los sujetos eran jóvenes emprendedores de Sierra Leona, país afectado por el ébola. Si bien los autores constatan la existencia de efectos positivos de la intervención, también señalan la heterogeneidad de los tamaños del impacto: los jóvenes menos preparados y más pobres canalizaron los beneficios del programa más hacia el consumo y menos hacia la creación de empresas que los jóvenes más preparados. Este hallazgo es valioso porque aporta a los responsables de las políticas información necesaria para dirigir con más precisión las PAE.

Por lo tanto, recomendamos a los responsables políticos que encarguen estimaciones de impacto a investigadores: a) que identifiquen claramente y por adelantado grupos de sujetos de los que pueden esperarse aspectos importantes e interesantes de heterogeneidad del impacto; y b) que financien experimentos de campo de un tamaño suficiente para que los datos puedan dividirse en submuestras que faciliten el examen detallado de los grupos de enfoque. Las muestras de gran tamaño también posibilitan que fuentes de heterogeneidad inesperadas pero

pertinentes para las políticas se examinen *ex post*, es decir, una vez concluido el experimento, posiblemente justificando experimentos de seguimiento que puedan revelar aspectos relevantes de la política.

Si por algún motivo no fuera factible llevar a cabo experimentos de campo, la heterogeneidad podría analizarse utilizando los demás métodos de los que se ha hablado en este artículo. El análisis con submuestras tampoco es la única herramienta de que dispone el investigador. Otra posibilidad es el modelo mixto, que permite que los parámetros de impacto del tratamiento como β_1 asuman distintos valores en distintos grupos. Y una alternativa más es la inclusión de efectos de interacción. Aún mejor es el análisis bayesiano, que incorpora heterogeneidad de los parámetros naturalmente en su marco de modelado. El análisis bayesiano tiene otras ventajas que describiré a continuación.

Recomendación 3: Realizar más análisis bayesianos

Retomando la idea anterior, es posible utilizar modelos jerárquicos bayesianos, estrategia ampliamente extendida para el examen de la heterogeneidad en un buen número de relaciones en el ámbito de la economía, el emprendimiento y administración de empresas (véanse, por ejemplo, Mackey *et al.*, 2017; Mackey y Dotson, 2024). Los modelos jerárquicos bayesianos van más allá del efecto mixto y las especificaciones de interacción, limitadas por estimarse solo las relaciones medias expresadas en forma de estimaciones puntuales. El análisis bayesiano, en cambio, representa gráficamente las distribuciones de parámetros y, en lugar de proporcionar p-valores para el contraste de hipótesis, informa de las probabilidades de que β_1 caiga en cualquier rango de valores determinado. Es fácil pensar en situaciones en las que esa información es más valiosa para los responsables de establecer las políticas que simplemente preguntar si los efectos “difieren significativamente de cero”, como ocurre con la inferencia estadística.

Por consiguiente, una ventaja clave del uso de modelos jerárquicos bayesianos para estudiar el impacto de las PAE es que la heterogeneidad se modela de forma natural. Una segunda ventaja es que

los modelos bayesianos pueden manejar muestras de pequeño tamaño, ya que la inferencia bayesiana se aplica a la muestra en lugar de recurrir al teorema del límite central para inferir los parámetros poblacionales. La tercera es que los métodos bayesianos incorporan explícitamente conocimientos previos para evaluar de qué modo un enfoque nuevo o nuevos datos modifican nuestro entendimiento de relaciones existentes. Así, miremos de nuevo la literatura relativa a subvenciones para la creación de nuevas empresas por demandantes de empleo que vimos en el apartado sobre subvenciones para la creación de nuevas empresas por demandantes de empleo. En lugar de que cada estudio trate sus estimaciones de forma independiente de los estudios previamente publicados, como es habitual, el análisis bayesiano nos permite incorporar lo que ya sabemos por estudios previos para refinar y perfeccionar nuestra base de conocimientos.

En pocas palabras, el uso de métodos bayesianos en el análisis de PAE sigue siendo un área de investigación desatendida, pero prometedora, que previsiblemente enriquecerá de un modo considerable nuestro entendimiento de los efectos de las PAE.

Recomendación 4: Prestar más atención a cuestiones de validez externa

En el caso de los experimentos, el análisis de intervenciones específicas suele conducir a validez interna alta, pero validez externa baja. Por ejemplo, si retomamos lo que decíamos en el apartado 2 de la sección tercera, podría ser útil aprovechar modificaciones legales históricas de programas de discriminación positiva para generar resultados relativos a un momento determinado del pasado de Estados Unidos; sin embargo, podría ser muy poco pertinente para las condiciones actuales de otros países.

Aunque podría parecer que los experimentos de campo son menos proclives a validez externa que los experimentos naturales, con frecuencia no es así. Para entender por qué, recordemos que los experimentos de campo que analizan programas de formación en gestión de empresas revisados en la sección cuarta generaban numerosas estimaciones diferentes de β_1 . Probablemente esto refleje

diferencias de contexto (país, período, etc.) de los experimentos, además de diferencias de contenido de los programas entre los estudios revisados. Por sí mismo, esto no es un problema, y de hecho probablemente es lo esperable. Sin embargo, puede utilizarse para animar a los responsables de las políticas a plantearse una pregunta más precisa que la genérica sobre si los programas de formación empresarial promueven el emprendimiento. Esa pregunta hace surgir cuestiones de validez externa. Supongamos, en cambio, que el responsable de formular la política preguntara: “¿Un programa de formación de tipo W promueve un resultado de desempeño de la función empresarial X entre un subgrupo de población Y en el país/contexto temporal Z?” Esta es una pregunta más adecuada porque tiene en cuenta explícitamente varias dimensiones contextuales existentes y reconoce la validez externa limitada de la que adolecen todos los estudios de impacto de las PAE.

De hecho, puede ser engañoso juzgar la eficacia de una política activa, y su posible idoneidad para ser adoptada en otros lugares, sin reconocer la cuestión de la validez externa. Por ejemplo, en la sección cuarta, se mencionó que la mayoría de los estudios del programa SUS alemán lo han asociado con impactos positivos para la posterior empleabilidad de personas desempleadas. El impacto de otras PAE de ese mismo tipo aplicadas en otros países ha sido más limitado en general (Parker, 2018). ¿Significa esto que esos otros programas son inferiores o que el contexto alemán es de algún modo especial? No necesariamente. El SUS iba dirigido a personas desempleadas de mayor capital humano (Caliendo *et al.*, 2015); otros programas, quizá atendiendo a loables motivos, podrían preferir dirigir la política activa a los trabajadores desempleados en situación de exclusión social. Esto bastaría para explicar los menores efectos de la política fuera de Alemania (Parker, 2018).

La lección en este caso es que el diseño del programa debe ser tenido en consideración y explicarse claramente. A veces, incluso lo que parece ser una política “simple” puede ser heterogénea y asumirse incorrectamente que algunos aspectos de esa heterogeneidad (por ejemplo, los destinatarios del programa) son un detalle irrelevante. Es nece-

sario reflexionar cuidadosamente para calibrar las expectativas políticas del impacto: qué es, quién lo recibe, cómo y cuándo.

En todos los casos, la fiabilidad y la validez interna pueden mejorarse si las estimaciones de impacto se consideran más en su totalidad y menos como entidades individuales. Concretamente, los estudios de replicación se están volviendo más habituales en la investigación sobre emprendimiento (por ejemplo, Seong y Parker, 2024). Al replicar estudios, podemos obtener una idea mejor de la solidez de hallazgos específicos. Así, la recomendación final será la siguiente:

Recomendación 5: Realizar más estudios de replicación de las PAE

VI. CONCLUSIÓN

Las PAE siguen siendo una herramienta relevante y popular utilizada por los Gobiernos y por distintos organismos del sector público de todo el mundo para promover la iniciativa empresarial. Dadas las ventajas sociales y económicas potenciales de la creación de empresas, es importante que los responsables de establecer las políticas estén al corriente de lo que ya sabemos y de cómo el análisis futuro puede mejorar esos conocimientos. Es probable que la “revolución causal” en el análisis de gestión acabe de empezar. Sería deseable que los responsables de establecer las políticas aumentaran su apoyo a una estimación rigurosa de los efectos de las PAE en el futuro basándose en las distintas lecciones esbozadas en este artículo. Entre otras, la necesidad de: insistir en diseños de investigación robustos y fiables, preferiblemente de asignación aleatoria de efectos del tratamiento; considerar atentamente las cuestiones de heterogeneidad y validez externa; y explotar el potencial de la estimación bayesiana, que por desgracia ha sido infrautilizada en la literatura existente.

NOTAS

(1) Véase **Parker (2018)**, capítulos 18-21, para obtener un panorama general de un amplio abanico de políticas y una revisión selectiva de la literatura. Véase también, el trabajo de **Parker (2024)** en el que se proporciona un marco para entender el modo en que la regulación afecta a la entrada y salida de emprendedores. **Cravo y Piza (2019)**, por su parte, realizan un metaanálisis de políticas relacionadas con los servicios de apoyo a pymes.

BIBLIOGRAFÍA

- Baumgartner, H. J. y Caliendo, M. (2008)**. Turning unemployment into self-employment: Effectiveness of two start-up programmes. *Oxford Bulletin of Economics and Statistics*, 70(3), 347-373.
- Berge, L. I. O., Bjorvatn, K. y Tungodden, B. (2015)**. Human and financial capital for microenterprise development: Evidence from a field and lab experiment. *Management Science*, 61(4), 707-722.
- Bertschek, I., Block, J., Kritikos, A. S. y Stiel, C. (2024)**. German financial state aid during Covid-19 pandemic: Higher impact among digitalized self-employed. *Entrepreneurship & Regional Development*, 36(1-2), 76-97.
- Block, J., Kritikos, A. S., Priem, M. y Stiel, C. (2022)**. Emergency-aid for self-employed in the Covid-19 pandemic: A flash in the pan? *Journal of Economic Psychology*, 93, 102567.
- Brooks, W., Donovan, K. y Johnson, T. R. (2018)**. Mentors or teachers? Microenterprise training in Kenya. *American Economic Journal: Applied Economics*, 10(4), 196-221.
- Bruhn, M., Karlan, D. y Schoar, A. (2018)**. The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico. *Journal of Political Economy*, 126(2), 635-687.
- Butler, I., Galassi, G. y Ruffo, H. (2016)**. Public funding for startups in Argentina: an impact evaluation. *Small Business Economics*, 46, 295-309.
- Caliendo, M. (2009)**. Star-up subsidies in East Germany: finally, a policy that works? *International Journal of Manpower*, 30(7), 625-647.
- Caliendo, M. y Kritikos, A. S. (2010)**. Start-ups by the unemployed: characteristics, survival and direct employment effects. *Small Business Economics*, 35(1), 71-92.
- Caliendo, M. y Künn, S. (2011)**. Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity. *Journal of Public Economics*, 95(3-4), 311-331.
- Caliendo, M. y Künn, S. (2015)**. Getting back into the labor market: The effects of start-up subsidies for unemployed females. *Journal of Population Economics*, 28, 1005-1043.
- Caliendo, M. y Künn, S. (2017)**. Regional effect heterogeneity of start-up subsidies for the unemployed. In *Entrepreneurship in a Regional Context* (170-196). Routledge.
- Caliendo, M., Künn, S. y Weißenberger, M. (2016)**. Personality traits and the evaluation of start-up subsidies. *European Economic Review*, 86, 87-108.
- Caliendo, M., Hogenacker, J., Künn, S. y Wießner, F. (2015)**. Subsidized start-ups out of unemployment: a comparison to regular business start-ups. *Small Business Economics*, 45, 165-190.
- Caliendo, M. y Tübbicke, S. (2020)**. New evidence on long-term effects of start-up subsidies: Matching estimates and their robustness. *Empirical Economics*, 59, 1605-1631.
- Chatterji, A. K., Chay, K. Y. y Fairlie, R. W. (2014)**. The impact of city contracting set-asides on black self-employment and employment. *Journal of Labor Economics*, 32(3), 507-561.

- Cho, Y. y Honorati, M. (2014).** Entrepreneurship programs in developing countries: A meta regression analysis. *Labour Economics*, 28, 110-130.
- Cravo, T. A. y Piza, C. (2019).** The impact of business-support services on firm performance: a meta-analysis. *Small Business Economics*, 53, 753-770.
- Drexler, A., Fischer, G y Schoar, A. (2014).** Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2), 1-31.
- Elert, N., Andersson, F. W. y Wennberg, K. (2015).** The impact of entrepreneurship education in high school on long-term entrepreneurial performance. *Journal of Economic Behavior & Organization*, 111, 209-223.
- Fairlie, R. W. (2023).** Evaluating entrepreneurship training: How important are field experiments for estimating impacts? *Journal of Economics & Management Strategy*, 32(3), 607-635.
- Fairlie, R. y Marion, J. (2012).** Affirmative action programs and business ownership among minorities and women. *Small Business Economics*, 39, 319-339.
- Fairlie, R. W., Karlan, D. y Zinman, J. (2015).** Behind the GATE experiment: Evidence on effects of and rationales for subsidized entrepreneurship training. *American Economic Journal: Economic Policy*, 7(2), 125-161.
- Gonzalez-Uribe, J. y Leatherbee, M. (2018).** The effects of business accelerators on venture performance: Evidence from start-up Chile. *The Review of Financial Studies*, 31(4), 1566-1603.
- Hinz, T. y Jungbauer-Gans, M. (1999).** Starting a business after unemployment: characteristics and chances of success (empirical evidence from a regional German labour market). *Entrepreneurship & Regional Development*, 11(4), 317-333.
- Hottenrott, H y Richstein, R. (2020).** Start-up subsidies: Does the policy instrument matter? *Research Policy*, 49(1), 103888.
- Karlan, D., & Valdivia, M. (2011).** Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics & Statistics*, 93(2), 510-527.
- Klinger, B. y Schündeln, M. (2011).** Can entrepreneurial activity be taught? Quasi-experimental evidence from Central America. *World Development*, 39(9), 1592-1610.
- Lafortune, J., Riutort, J. y Tessada, J. (2018).** Role models or individual consulting: The impact of personalizing micro-entrepreneurship training. *American Economic Journal: Applied Economics*, 10(4), 222-245.
- Lee, D. S. y Lemieux, T. (2010).** Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281-355.
- Mackey, T. B. y Dotson, J. P. (2024).** Bayesian Statistics in Management Research: Theory, Applications, and Opportunities. In *Oxford Research Encyclopedia of Business and Management*.
- Mackey, T. B., Barney, J. B. y Dotson, J. P. (2017).** Corporate diversification and the value of individual firms: A Bayesian approach. *Strategic Management Journal*, 38(2), 322-341.
- Mano, Y., Iddrisu, A., Yoshino, Y. y Sonobe, T. (2012).** How can micro and small enterprises in Sub-Saharan Africa become more productive? The impacts of experimental basic managerial training. *World Development*, 40(3), 458-468.
- McKenzie, D. y Woodruff, C. (2014).** What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer*, 29(1), 48-82.
- OECD (2023).** *The Missing Entrepreneurs 2023. Policies for Inclusive Entrepreneurship and Self-employment*, OECD, Paris. https://www.oecd.org/en/publications/the-missing-entrepreneurs-2023_230efc78-en.html
- Oosterbeek, H., Van Praag, M. y Ijsselstein, A. (2010).** The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3), 442-454.
- Parker, S. C. (2018).** *The Economics of Entrepreneurship*. 2nd Edition. Cambridge, England: Cambridge University Press.

- Parker, S. C. (2024).** Public policy impacts on entrepreneurial entry and exit: a real options framework. *Academy of Management Perspectives*, in press.
- Pfeiffer, F. y Reize, F. (2000).** Business start-ups by the unemployed—an econometric analysis based on firm data. *Labour Economics*, 7(5), 629-663.
- Premand, P., Brodmann, S., Almeida, R., Grun, R. y Barouni, M. (2016).** Entrepreneurship education and entry into self-employment among university graduates. *World Development*, 77, 311-327.
- Quinn, S. y Woodruff, C. (2019).** Experiments and entrepreneurship in developing countries. *Annual Review of Economics*, 11(1), 225-248.
- Roth, J., Sant’Anna, P. H., Bilinski, A. y Poe, J. (2023).** What’s trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2), 2218-2244.
- Seong, M. y Parker, S. C. (2024).** Does gendered wording in job advertisements deter women from joining start-ups? A replication and extension of Gaucher, Friesen, and Kay (2011). *Strategic Entrepreneurship Journal*, 18(2), 286-305.
- Storey, D. J. (2000).** Six steps to heaven. *Handbook of Entrepreneurship*. Oxford: Blackwells.
- Von Graevenitz, G., Harhoff, D. y Weber, R. (2010).** The effects of entrepreneurship education. *Journal of Economic Behavior & Organization*, 76(1), 90-112.
- Rosas, N., Acevedo, M. C. y Zaldivar, S. (2022).** Starting points matter: Cash plus training effects on youth entrepreneurship, skills, and resilience during an epidemic. *World Development*, 149, 105698.
- Wolff, J., Nivorozhkin, A. y Bernhard, S. (2016).** You can go your own way! The long-term effectiveness of a self-employment programme for welfare recipients in Germany. *International Journal of Social Welfare*, 25(2), 136-148.