



Un passeig per les tècniques d'avaluació d'impacte

Marcos Vera Hernández és professor lector del Departament d'Economia del University College London i investigador associat de l'Institute for Fiscal Studies. Ha treballat en avaluacions d'impacte a Colòmbia, Hondures, Jamaica, Malawi, l'Índia i Catalunya. Ha impartit cursos en escoles d'estiu i tallers sobre avaluació d'impacte, i també ha escrit algunes notes introductòries i revisat llibres dedicats a l'avaluació d'impacte.

Què és una avaluació d'impacte i com es fa? Segurament aquestes preguntes ja se les han fetes moltes persones, atès l'augment de la popularitat d'aquest tipus d'avaluacions en els últims anys. Com que no hi ha avaluació sense programa, i no hi ha programa sense acrònim, ens n'inventarem un: el POJ o Programa per a l'ocupabilitat dels joves – programa insígnia del PSI (Partit Sense Ideologia), que va començar l'1 de gener de 2011 i pretén disminuir la taxa d'atur juvenil a través de pràctiques en empreses de 6 mesos de durada (fins al 30 de juny de 2011). Però no deixeu de llegir si treballeu en un sector diferent del laboral (salut, educació, etc.), ja que amb el 95% de confiança podreu aplicar el que tractarem en aquest article a altres sectors de l'Administració pública diferent del laboral.

L'avaluació d'impacte del POJ consistiria a calcular quin és el valor afegit del POJ en les variables de resultat que considerem importants; per exemple, la taxa d'ocupació juvenil després d'un any de finalitzar el programa, és a dir, el 30 de juny de 2012.

Per calcular aquest valor afegit necessitem calcular la taxa d'ocupació dels joves que van participar en el POJ, a més de calcular quina hauria estat la taxa d'ocupació a 30 de juny de 2012 d'aquests mateixos joves participants en el POJ si no hi haguessin participat. La diferència d'aquestes dues taxes ens dona el valor afegit, o impacte, del programa en termes d'ocupació.

La taxa d'ocupació dels participants en el POJ és fàcil de calcular, n'hi hauria prou amb fer una enquesta als participants i preguntar-los si treballen, o fins i tot es podrien fer servir registres administratius. El que és difícil és calcular la taxa d'ocupació que els participants haurien tingut si no hi haguessin participat. I és difícil perquè consisteix a esbrinar què hauria passat a un grup de persones en una situació irreal: els joves van participar en el POJ; per tant, calcular què els hauria passat si no hi haguessin participat necessitarà, almenys, una certa reflexió. Hom podria pensar a cridar els participants en el POJ i preguntar-los si creuen que treballarien



si no haguessin participat en el programa, però segurament les respostes no serien fiables. També hem de resistir la temptació de pensar que si els participants no hi haguessin participat, la seva taxa d'ocupació seria zero. Fins i tot en temps difícils, hi haurà joves que haurien aconseguit feina fins i tot si el POJ no hagués existit. També seria molt arriscat suposar que la taxa d'ocupació dels participants del POJ seria la mateixa que la de la mitjana de la població jove, la qual es podria calcular utilitzant dades administratives. Els participants en el POJ els van triar les empreses per fer les pràctiques, per la qual cosa segurament tenen qualitats especials (motivació, interès, bona presència, millors connexions, etc.) que faciliten que trobin feina, fins i tot si no haguessin participat en el POJ. Per tant, la taxa mitjana de la població jove no seria un bon indicador de la taxa mitjana d'ocupació dels participants en el POJ si no hi haguessin participat.

No cregueu que dir que els participants són diferents dels no participants sigui una curiositat acadèmica. Cap al 2001, abans de deixar Barcelona per anar-me'n a Londres, em vaig entrevistar amb un alt càrrec d'un ens públic associat a la Generalitat perquè estaven implementant un programa que a mi m'interessava molt avaluar. Quan li vaig pre-

guntar com havien triat els participants, em va contestar que havien escollit els millors perquè volien que el programa semblés reeixit. I m'ho va dir sense cap mena de vergonya!

Així doncs, el que intentarem descriure és com podem dissenyar les avaluacions per obtenir estimacions d'impacte (valor afegit) fiables. El mètode que, en principi, sol plantejar menys crítiques metodològiques és el **d'elecció aleatòria de participants**. Aquest mètode consisteix a triar a l'atzar els participants del POJ entre tots els que tenen els requisits per ser-ne beneficiaris i a més hi han volgut participar. Per tant, exclourà durant un cert període de temps un grup d'individus que han volgut participar en el POJ. Però, en el fons, aquesta situació d'«excés de demanda» sol ser bastant habitual, atès que mai no hi ha prou recursos perquè hi participin tots els que ho volen. La diferència amb el que fem habitualment és que, en comptes de deixar carta blanca a les empreses de formació perquè triïn els participants, o fer l'elecció mitjançant un sistema rígid (i fins a cert punt arbitrari) de punts dissenyat en un despatx, s'opta per fer l'elecció de manera aleatòria.

Però, a què es deu la transparència i altres



fortaleses del mètode d'elecció aleatòria? Com ja hem avançat, el que resulta difícil de l'avaluació d'impacte del POJ és estimar la taxa d'ocupació dels participants si no hi haguessin participat. En el cas de l'elecció aleatòria, el grup de no participants té les mateixes característiques que el grup de participants, ja que l'única diferència entre ells és que uns van treure cara i els altres van treure creu en el sorteig en el qual es va decidir la participació. En altres paraules, els no participants són idèntics als participants, amb l'única diferència que no van participar en el POJ. Per això, resulten adequats per mesurar quina hauria estat la taxa d'ocupació dels participants si no hi haguessin participat, la qual cosa, com hem dit, és la clau de l'avaluació d'impacte. Després de restar aquesta quantitat a la taxa d'ocupació dels participants en el POJ, obtenim el valor afegit o impacte del programa.

Exemple de selecció aleatòria de participants

Training Disadvantaged Youth in Latin America: Evidence from a Randomized Trial

Què: Avaluen l'impacte d'un programa de capaciació de joves. El programa combina classes amb feina en empreses.

On: Colòmbia

Resultats: El programa augmenta els ingressos i l'ocupació d'homes i dones, però l'efecte és més gran per a les dones.

Autors: O. Attanasio, A. Kugler, C. Meghir

Publicat a: *American Economic Journal: Applied Economics* (2011), 3: 188-220.

De vegades no hi ha voluntat política per implementar un disseny d'avaluació basat en el model d'elecció aleatòria de participants. Quan passa això, hi ha dissenys que continuen conservant un element d'elecció aleatòria però que resulten políticament més fàcils d'implementar. El primer mètode consisteix a modificar el mètode d'elecció aleatòria permetent que els individus inicialment assignats al grup de no participants puguin ser reassignats al grup de participants si apelen a l'entitat encarregada de gestionar el POJ. El millor per a l'avaluació és dissenyar un sistema d'apel·lació que minimitzi el nombre d'apel·lacions perquè la diferència amb l'elecció aleatòria inicial sigui tan petita com sigui possible. El segon mètode consisteix a donar una «empenteta» a un grup d'individus triats a l'atzar perquè la seva participació en el POJ sigui més probable que la del grup al qual no vam donar l'«empenteta». Aquesta «empenteta» pot ser un incentiu econòmic, cobrir les despeses de transport, o fins i tot trucar-li per telèfon per recordar-li que pot participar en el POJ si ho vol. Aquesta modalitat de disseny aleatori és molt útil en programes als quals no es pot negar l'accés a ningú, és a dir, programes d'accés universal.

Ara descriurem les tècniques d'avaluació



d'impacte més rellevants quan no s'ha pogut fer l'elecció aleatòria de participants (o alguna de les variants descrites prèviament). La primera tècnica que descriurem s'anomena **diferències en diferències**, el significat de la qual es pot entendre fent servir la dita de «la unió fa la força», malgrat que potser és més apropiat inventar-nos-en una *ad hoc* i dir que la combinació ens acostava a la veritat. Si entrem en detalls, el mètode de **diferències en diferències** ens proporciona l'estimació de l'impacte combinant dues estratègies que totes soles serien insuficients. La primera estratègia (insuficient tota sola) consisteix a restar la taxa d'ocupació dels participants en el POJ abans que el POJ comencés (1 de gener de 2011) a la taxa d'ocupació d'aquells mateixos participants un cert temps després que s'acabés la seva participació en el programa (30 de juny de 2012). En unes altres paraules, la primera estratègia insuficient és la diferència entre l'*abans* i el *després*. Resulta insuficient perquè les condicions macroeconòmiques poden ser força diferents entre l'*abans* i el *després*, i això es reflectirà en la diferència de les taxes d'ocupació. Per exemple, pot ser que el POJ sigui un programa molt bo, però que les taxes d'ocupació a 30 de juny de 2012 siguin pitjors que les taxes a 1 de gener de 2011, simplement perquè

la crisi s'hagi aguditzat (per la qual cosa és possible que, en absència del POJ, les taxes d'ocupació haurien estat fins i tot pitjors). Abans d'acabar amb l'estratègia de comparar l'*abans* i el *després*, cal destacar que el seu gran avantatge és que només exigeix utilitzar dades de participants, per la qual cosa no tenim el problema de comparar participants i no participants que poden tenir característiques diferents que els fan més o menys atractius per als ocupadors (motivació, intel·ligència, etc.). Això ens porta a la segona estratègia que resulta insuficient tota sola, comparar la taxa d'ocupació el 30 de juny de 2012 entre participants i no participants. Com ja hem anticipat, la raó per la qual resulta insuficient és que participants i no participants poden tenir diferents característiques personals, i per tant la taxa d'ocupació a 30 de juny de 2012 no sols es veurà afectada per la participació en el POJ, sinó també per les esmentades característiques, que, com que poden ser diferents entre participants i no participants, poden contaminar la comparació. No obstant això, el seu avantatge és que la comparació es fa en una mateixa data (30 de juny de 2012), per la qual cosa les condicions macroeconòmiques són les mateixes per a participants i no participants.



Resumint, l'estratègia de comparar l'*abans* amb el *després* té l'avantatge que només s'utilitzen participants, per la qual cosa no hi ha el problema de comparar persones amb unes característiques diferents, però té el desavantatge que les condicions macroeconòmiques poden ser diferents entre l'*abans* i el *després*. Amb l'estratègia de comparar la taxa d'ocupació de *participants* i *no participants* passa gairebé el contrari: la diferència de condicions macroeconòmiques no és un problema, però el fet que comparem persones amb unes característiques diferents sí que ho és. Per tant, no és estrany que, quan es combinen aquestes dues estratègies, es pugui treure profit de les virtuts i contrarestar els desavantatges de cada estratègia per separat. I per això dic que el mètode de **diferències en diferències** ens proporciona l'estimació de l'impacte combinant dues estratègies que totes soles serien insuficients. Malgrat que aquest mètode pugui donar estimacions fiables de l'impacte, s'ha de destacar que es necessita que es compleixin certs supòsits perquè les estimacions siguin de confiança. En particular, és necessari assumir que la taxa d'ocupació entre participants i no participants evolucionaria en el temps de manera paral·lela si no existís el POJ.

Exemple de diferències en diferències

Providing employers with incentives to train low-skilled workers: evidence from the U.K. employer training pilots

Què: Avaluen l'impacte d'un programa que donava incentius als empresaris per capacitar els empleats de baixa qualificació.

On: Regne Unit

Resultats: El programa no va augmentar no va augmentar el nivell de capacitació dels empleats de baixa qualificació.

Autors: L. Abramovsky, E. Battistin, E. Fitzsimons, A. Goodman, H. Simpson

També voldria esmentar el mètode d'**aparellament** (*matching*), la popularitat del qual ha evolucionat una mica a l'estil de muntanya russa (la meua impressió és que ara torna a estar en un punt baix, però ha arribat a estar en un punt bastant alt). El mètode d'aparellament consisteix a assumir que l'analista té dades sobre totes les característiques individuals, familiars i d'ambient que afecten tant la taxa d'ocupació com la decisió de participar en el POJ. El mètode consisteix a donar ponderacions als individus no participants perquè les característiques personals, familiars i d'ambient d'aquest grup ponderat siguin les mateixes que les del grup de participants. Per exemple, si el grup de participants té un nivell educatiu mitjà més alt que el grup de no participants, llavors el mètode d'aparellament donarà una ponderació més alta als individus més educats que



formin part del conjunt d'individus no participants, amb l'objectiu que l'educació mitjana sigui la mateixa en el grup de participants que en el grup ponderat de no participants. Això permet comparar la taxa d'ocupació dels participants amb la taxa d'ocupació dels no participants, utilitzant les ponderacions que els fan el més semblants possible als participants.

La raó per la qual (crec) el mètode d'aparellament ha baixat en popularitat és perquè el supòsit necessari perquè funcioni (que l'analista té dades sobre totes les característiques individuals, familiars i d'ambient que afecten la taxa d'ocupació i la decisió de participar en el programa) es considera massa restrictiu, almenys en alguns casos, en què variables com ara motivació, intel·ligència, etc. són molt difícils de mesurar i, per tant, l'analista no en sol tenir dades. De tota manera, alguns investigadors han indicat que utilitzar dades de la variable de resultat (taxa d'atur en el cas que ens ocupa), però per a diversos períodes anteriors al començament del programa, pot ajudar molt a obtenir resultats fiables.

L'última tècnica objecte d'aquesta breu revisió és la de **regressió discontinua**. S'aplica

Exemple del mètode d'aparellament (*matching*)

Long-Run Effects of Public Sector Sponsored Training in West Germany

Què: Avaluen l'impacte a llarg termini de diversos programes de capacitació per a desocupats.

On: Alemanya de l'Oest

Resultats: La participació en programes de capacitació augmenta considerablement la taxa d'ocupació a llarg termini.

Autors: M. Lechner, R. Miquel, C. Wunsch

Publicat a: *Journal of the European Economic Association* (2011), 9: 742-784.

quan la participació en el programa depèn del fet que una variable prengui un conjunt de valors determinats. Per exemple, podríem escriure en les normes del POJ que només els individus amb una renda familiar menor a un cert punt de tall (per exemple, 31.256 euros) hi poden participar. Una altra alternativa podria ser utilitzar un sistema de punts (per exemple, els individus obtenen punts en funció de la mida de la llar, el nivell educatiu del cap de família, l'edat, etc.) i permetre que hi participin només els que tenen més punts que un cert valor predeterminat fixat en les bases del POJ. En qualsevol cas, a l'efecte de la implementació de la tècnica, no importa si el punt de tall (de la renda familiar o del sistema de punts) és més gran o més petit, tan sols importa que hi hagi un punt de tall.

La lògica del mètode de **regressió discontinua** és la següent. El fet que la participació



sigui determinada per un punt de tall comportarà que hi hagi un salt, justament al llinyar, en el percentatge d'individus que hi participin. Així, en l'exemple anterior, esperaríem que el percentatge d'individus que participen en el POJ amb una renda familiar de 31.255 euros fos molt més gran que per a individus amb una renda familiar de 31.257 euros. Per tant, esperem un salt en el percentatge de participació en el punt de tall de 31.256 euros. Doncs bé, si el POJ resulta efectiu per augmentar la taxa d'ocupació, aleshores també esperaríem que el salt en el percentatge de participació es reflectís en la taxa d'ocupació i, per tant, hi hagués un salt entre la taxa d'ocupació d'individus amb una renda familiar de 31.255 euros i de 31.257 euros.

Perquè el mètode de **regressió discontinua** funcioni és necessari suposar que no hi ha cap altra variable que «salti» en el mateix punt de tall. Per exemple, no podríem aplicar el mètode de regressió discontinua si el punt de tall de 31.256 euros també coincidís amb el punt de tall d'algun altre programa o política (cursos de formació, pensió no contributiva, etc.), ja que si fos així, no podríem saber quina part del «salt» en la taxa d'ocupació és deguda al POJ i quina part a l'altre programa

que també mostra un salt. Una altra condició important és que els individus no puguin triar en quin costat del punt de tall es posen. Per exemple, l'estimació quedaria invalidada si els individus més motivats poguessin reduir la renda de 31.257 euros a 31.255 euros amb l'objectiu de poder participar en el POJ. En aquest cas, no podríem saber quina part del salt en la taxa d'ocupació és deguda al POJ, i quina a la diferent motivació que tenen els individus la renda familiar dels quals és de 31.257 euros comparada amb la dels individus la renda familiar dels quals és de 31.255 euros.

Una cosa important que s'ha de tenir en compte és que el mètode de regressió discontinua tan sols informa sobre l'impacte per a individus que estan prop del punt de tall. Per exemple, encara que es compleixin els supòsits del paràgraf anterior, el mètode no proporcionaria estimacions vàlides sobre l'impacte del programa en el cas d'individus la renda familiar dels quals fos de 20.000 euros, per exemple, atès que estan lluny del punt de tall de 31.256 euros.

Recapitulant, després de definir l'impacte en termes de valor afegit d'un programa, hem descrit el mètode d'elecció aleatòria i les se-



Exemple de regressió discontinua

The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates over Twenty Years

Què: Avaluen l'efecte de la durada de la prestació d'atur en la durada de l'atur.

On: Alemanya

Resultats: L'efecte és significatiu, però té una magnitud reduïda. L'efecte no canvia amb el cicle econòmic.

Autors: J. Schmieder, Till von Wachter, Stefan Bender

Publicat a: *Quarterly Journal of Economics* (en procés de publicació)

ves variants, i també mètodes que no necessiten l'elecció aleatòria de participants. Respecte a aquests tres últims (diferències en diferències, aparellament i regressió discontinua), hem destacat que la validesa dels mètodes descansa en uns determinats supòsits que no necessàriament s'han de complir en la realitat. És per això que s'ha de fer un acurat treball empíric per poder avaluar la plausibilitat d'aquests supòsits, i així informar sobre la validesa de les estimacions d'impacte.

Per saber-ne més.

A continuació suggereixo una sèrie de lectures per a qui vulgui aprofundir en el tema. Les lectures (1) i (2) són molt breus però són fàcils de llegir (bé, això espero), i haurien d'ajudar a aclarir els conceptes més bàsics.

Les lectures (3) i (4) són llibres excel·lents sobre el tema de l'avaluació d'impacte. Tots dos utilitzen molts exemples, i estan molt ben documentats, fins i tot contenen les dades i els codis de programes d'ordinador per emprar les tècniques descrites. La lectura (5) és una revisió dels últims avenços metodològics, i la (6) és una discussió molt interessant sobre les tècniques d'elecció aleatòria de participants. La lectura (2) també conté una bibliografia comentada més extensa que la que presentem aquí.

(1) Ravallion, M. «The mystery of the vanishing benefits. Ms. Speedy Analyst's introduction to evaluation». *World Bank Economic Review* (2001), 15: 115-140.

(2) Fitzsimons, E.; Vera-Hernández, M. «A practitioner's guide to evaluating the impacts of labor market programs». *World Bank Employment Policy Primer* (December 2009).

<http://goo.gl/bzpGg>

(3) Bernal, R.; Peña, X. *Guía práctica para la evaluación de impacto*. Bogotá: Publicaciones CEDE. Universidad de Los Andes, 2011.

<http://goo.gl/k6QjT>

(4) Gertler, P. [et al.] *La evaluación de impacto en la práctica*. Washington DC: World Bank Training Series, 2011. <http://goo.gl/uKfs1>

(5) Imbens, G. W.; Wooldridge, J. «Recent Developments in the Econometrics of Program Evaluation». *Journal of Economic Literature* (2009), 47: 5-86.

(6) Burtless, G. «The case for randomized field trials in economic and policy research». *Journal of Economic Perspectives* (1995), 9: 63-84.

