

Políticas de mercado de trabajo basadas en datos objetivos - Fundamentos metodológicos e innovadores Ejemplo

Profesor Dr. Marco Caliendo
(Universidad de Potsdam y Oficina de evaluación)

`caliendo@uni-potsdam.de`

`www.caliendo.de`

`www.evaluation-office.de`

Congreso Mundial de la AMSPE
28 de abril de 2022

Evaluación del impacto contrafactual

- **Objetivo:** Calcular el impacto causal de una determinada política en las «unidades» afectadas.
 - Pregunta fundamental a la que responder:
«¿Qué habría pasado si las unidades afectadas no hubieran recibido el tratamiento?» (resultado contrafactual)
- **Efecto causal:** Comparación de los resultados observados con la situación **contrafactual**.
- **Problema fundamental de la evaluación:** este supuesto contrafáctico no se observa nunca (para la misma unidad en el mismo momento).
⇒ Por lo tanto, debemos encontrar **una buena muestra** (proxy) en un grupo de control.

Sesgo de selección

- El principal problema al que nos enfrentamos con los «enfoques estándar» (p. ej., las comparaciones antes y después/transversales) es que la asignación al grupo de tratamiento y al grupo de control **no es aleatoria**.
- Los participantes y los no participantes pueden **diferir incluso en ausencia del programa**.



Grupo de tratamiento



Grupo de control

- Por lo tanto, una simple comparación (promedio) no es significativa debido al **sesgo de selección**.

Resolver el problema de la selección

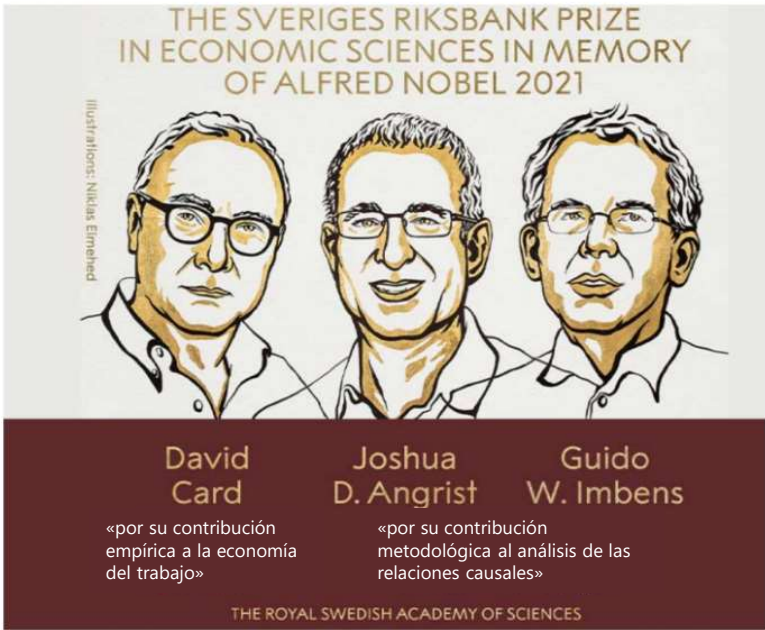
– Hay una variedad de métodos consolidados para superar el sesgo de selección.
Nuestro enfoque en la actualidad:

Métodos experimentales



El Premio Sveriges Riskbank 2019 en Ciencias Económicas en memoria de Alfred Nobel se ha otorgado conjuntamente a Abhijit Banerjee, Esther Duflo y Michael Kremer «por su enfoque experimental de la reducción de la pobreza mundial».

Métodos cuasi/no experimentales



Índice

- 1 **Introducción**
- 2 **Marco de evaluación**
- 3 **Métodos de evaluación**
 - 1 **Experimentos**
 - 2 **Adecuación**
 - 3 **Diferencia en diferencias (DID)**
 - 4 **Diseño de regresión discontinua (RDD)**
- 4 **Conclusión**

Evaluación del programa - Supuesto en las mejores condiciones posibles (2)

- En las mejores condiciones posibles, el evaluador ya participa en las primeras fases del diseño del programa y tiene influencia en los datos recogidos para una evaluación posterior.
- Estas fases son:
 1. Definir los objetivos del programa
 - Desarrollar una teoría del cambio
 - Diseñar el programa
 - Aplicar y recoger datos de referencia
 - Recopilar datos sobre los resultados finales
 - Evaluación de impacto contrafactual
- La evaluación de los procesos (centrada en la ejecución y el funcionamiento del programa) y la evaluación del impacto deben considerarse complementarias.
- Podemos utilizar la información obtenida en la evaluación de los procesos para seleccionar otros criterios de evaluación.

Evaluación del programa - Supuesto en las mejores condiciones posibles (2)

Preguntas importantes a las que se debe dar respuesta ya en la fase de diseño:

– **Objetivos y medición del éxito:**

- ¿Cuáles son los efectos previstos del programa?
- ¿Cómo se mide el éxito del programa?

– **Teoría del cambio:**

- ¿Cuál es la secuencia de acontecimientos que conduce a los resultados observados?
- ¿Cuáles son los diferentes parámetros que contribuyen al éxito del programa?

– **Estrategia empírica:**

- ¿Qué tipo de metodología de evaluación debe seguirse?
- ¿Cómo se recopilarán los datos necesarios?
- ¿Cómo distinguir qué mecanismos teóricos son los más importantes?

⇒ En las mejores condiciones posibles, los evaluadores disponen de suficiente **tiempo, presupuesto y datos de alta calidad.**

Evaluación del programa - Supuesto en las condiciones reales

– Sin embargo, en **condiciones reales**, las evaluaciones se llevan a cabo a menudo en circunstancias que no son óptimas («**evaluaciones en último término**»):

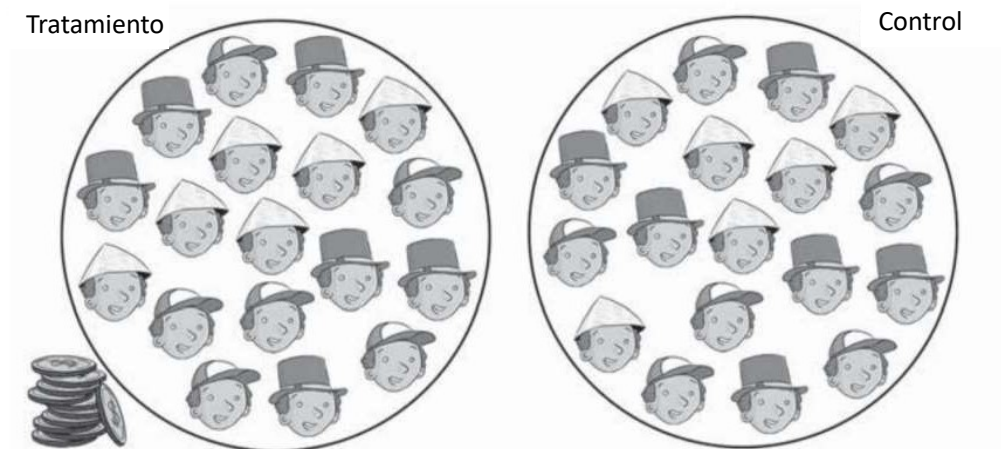
Limitaciones con las que deben realizarse las evaluaciones

Tiempo	Presupuesto	Datos	Condiciones típicas
X			Se recurre tarde al evaluador y con un plazo limitado
	X		Dificultades para recopilar los datos del estudio
		X	No se dispone de datos de referencia, es un tema delicado y la recopilación de datos es difícil
X	X		Se dispone de datos secundarios disponibles, pero de poco tiempo para analizarlos
		X	Poco tiempo y no se recogieron datos
	X	X	diseño limitado del estudio por falta de tiempo
			Se llama tarde al evaluador, el plazo no es un problema
			No hay acceso a los datos básicos, el presupuesto es ajustado
X	X	X	Se llama tarde al evaluador, con un plazo y un presupuesto ajustados, sin datos de referencia y sin grupo de control definido

Fuente: Bamberger et. al. (2004)

Experimentos

- **Los experimentos** asignan unidades de la población elegible de **forma aleatoria**.
- Esto garantiza que la participación no esté relacionada con las características de las unidades y equilibra las muestras en características observadas y no observadas.



Source: Evaluation in Practice

- Por lo tanto, las diferencias observadas en los resultados entre los dos grupos pueden atribuirse **únicamente** al tratamiento.
- **Estimador**: Diferencias medias transversales simples en el resultado Y.

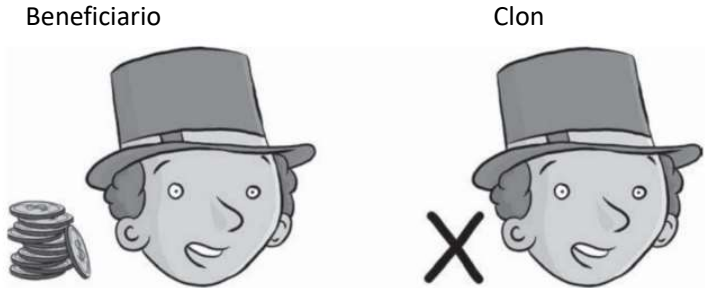
Ejemplo de experimento: Transferencia de puestos de trabajo

Crépon *et al.* (2013) ¿Tienen las políticas del mercado laboral efectos de transferencia? Datos de un experimento aleatorio por grupos

- **Interrogante investigado:** ¿Cuál es el efecto directo e indirecto (transferencia) de la ayuda a la inserción laboral en los resultados del mercado de trabajo de los jóvenes demandantes de empleo formados en Francia?
- **Tratamiento:** Participación en un programa de inserción laboral.
- **Datos:** Datos administrativos y de estudios, recopilados entre 2007 y 2010.
 $N \approx 22\,000$.
- **Método:** Experimento de campo controlado aleatorio. Los jóvenes desempleados y con estudios fueron asignados aleatoriamente a un grupo de tratamiento o de control.
- **Resultados:** Hay un efecto positivo para los jóvenes tratados sobre la búsqueda de empleo, mientras que los jóvenes no tratados tienen menos probabilidades de encontrar un trabajo (efecto de transferencia).

Adecuación (Matching)

- **Idea:** Para cada participante, elija uno (o varios) **gemelos estadísticos** de la muestra de no participantes.
- Deben ser idénticos en **todas las características relevantes**. Se trata de una exigencia muy fuerte que requiere datos informativos.



Source: Evaluation in Practice

- Al igual que en un experimento, esto da lugar a una muestra equilibrada.
- **Estimador:** Diferencias de las medias transversales simples en el resultado de **la muestra emparejada**.
- **La maldición de la dimensionalidad:** Si el número de características relevantes es grande, puede ser muy difícil encontrar una **correspondencia exacta**.
- **Una solución:** el emparejamiento por « tendencia-puntuación » resume toda la información en un solo índice y selecciona al no participante más cercano con respecto a ese índice.

Unos datos mejores ayudan mucho.

- Aplicar un enfoque de adecuación creíble no es fácil.

Unos datos mejores ayudan mucho.

- A menudo, las estimaciones pueden mejorarse combinando varias fuentes de datos:
 - Los datos a nivel individual y de empresa suelen estar disponibles a bajo coste en los registros administrativos (p. ej., a través de las agencias de empleo).
 - Los datos regionales/nacionales los proporcionan los datos (inter)nacionales de **las agencias de estadística**.
- **Ejemplos:**
 - El conjunto de datos de evaluación de IZA combina datos administrativos y de encuestas que permiten enriquecer los datos de la administración con información sobre «características normalmente no observadas»
 - DellaVigna *et al.* (2022) realizan una encuesta por SMS dos veces por semana durante cuatro meses.

Ejemplo de adecuación: Subvenciones para la puesta en marcha

Caliendo/Künn/Weißenberger (2016): Rasgos de personalidad y subvenciones para la puesta en marcha

- **Interrogante investigado:** ¿Son las subvenciones para la puesta en marcha para desempleados un programa activo eficaz en el mercado laboral? ¿Y los rasgos de personalidad omitidos suponen una amenaza para la fiabilidad de las estimaciones correspondientes?
- **Tratamiento:** Los desempleados dispuestos a crear una empresa reciben transferencias mensuales durante un máximo de 15 meses.
- **Datos:** Combinación de datos administrativos y de estudios. $N \approx 1\,300$.
- **Origen del sesgo de selección:** Los participantes se autoseleccionan en el programa; las características de los participantes difieren de las de los no participantes.
- **Método:** Adecuación de participantes y no participantes en función de una amplia gama de características y resultados previos al tratamiento.
- **Resultados:**
 - Efectos positivos sobre las probabilidades de empleo y los ingresos.
 - Los resultados se refuerzan tomando en consideración rasgos de personalidad que por lo general no se tienen en cuenta.

Diferencia en diferencias (1)

- Las configuraciones de **diferencia en diferencias** (DID) suelen utilizar una especie de «experimento natural» que se produce debido a un cambio de política, en el que un grupo de unidades se ve afectado por el tratamiento mientras que el otro grupo no.
 - **Por ejemplo:** un Estado aumenta el salario mínimo, pero el estado vecino no lo hace.
- **Importante:** La DID asume tendencias temporales paralelas (TTP) para el grupo de tratamiento y el de control sin tratamiento y permite diferentes niveles de pretratamiento («sesgo de base»).
- **Validez de las TTP:**
 - La inspección de la similitud de los patrones de pretratamiento proporciona algunas indicaciones de la probabilidad de que la hipótesis de las TTP sea válida.
 - Unas tendencias de pretratamiento significativamente diferentes arrojan serias dudas sobre la fiabilidad de las estimaciones.

Diferencia en diferencias (2)

- **Intuición del estimador de DID:** Combinar las estimaciones antes y después de los grupos de tratamiento y control.
 - Al comparar los cambios dentro de los grupos, controlamos implícitamente los **factores no observados constantes a lo largo del tiempo.**
 - Al comparar estos cambios entre los grupos, también controlamos las **tendencias temporales de los resultados.**
- **Estimador:** Comparar los cambios en el grupo de tratamiento con los cambios en el grupo de control a lo largo del tiempo.

Ejemplo de DID - Salario mínimo (1)

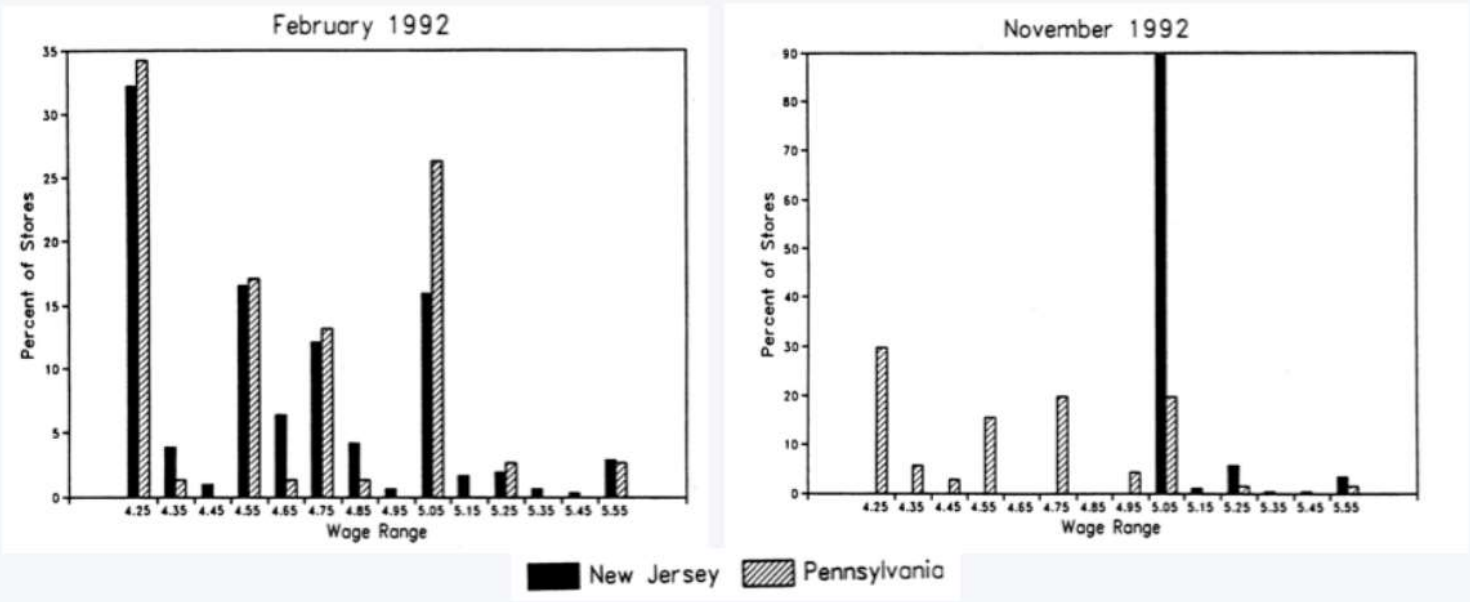
Card/Krüger (1994): Salario mínimo y empleo: un estudio de caso del sector de la comida rápida en Nueva Jersey y Pensilvania

- **Interrogante investigado:** ¿Impacto del aumento del salario mínimo en el empleo de salario bajo?
- **Tratamiento:** El salario mínimo aumentó de 4,25 a 5,05 dólares por hora en Nueva Jersey en abril de 1992.
- **Datos:** Datos de la encuesta sobre salarios y empleo para N = 410 restaurantes de comida rápida de Nueva Jersey y Pensilvania.
- **Origen del sesgo de selección:** Los restaurantes no afectados de Nueva Jersey pueden atender a otros clientes y ofrecer comidas más caras.
- **Método:** Comparar la evolución del empleo a tiempo completo en los restaurantes de comida rápida de Nueva Jersey y del estado vecino, PA.

Ejemplo de DID - Salario mínimo (2)

Card/Krüger (1994): Salario mínimo y empleo: un estudio de caso del sector de la comida rápida en Nueva Jersey y Pensilvania

– Comparación descriptiva de los salarios antes y después del tratamiento.



Ejemplo de DID - Salario mínimo (2)

Card/Krüger (1994): Salario mínimo y empleo: un estudio de caso del sector de la comida rápida en Nueva Jersey y Pensilvania

– Cálculo de los rendimientos medios de la muestra (s.e. entre paréntesis):

Variable	Estruc. por Estado		
	PA (i)	NJ (ii)	Diferencia, NJ – PA (iii)
1. ETC empleados antes, todas las observaciones disponibles	23,33 (1,35)	20,44 (0,51)	-2,89 (1,44)
2. ETC empleados después, todas las observaciones disponibles	21,17 (0,94)	21,03 (0,52)	-0,14 (1,07)
3. Variación del empleo medio en los ETC	-2,16 (1,25)	0,59 (0,54)	2,76 (1,36)

Source: Card/Krueger (1994), p. 780

Diseño de regresión discontinua (RDD)

- Muchos programas funcionan con un determinado **umbral de admisibilidad** para ciertos índices.
- Por ejemplo, los trabajadores que superan una determinada edad reciben **prestaciones por desempleo** durante un periodo más largo.
- El RDD (bruto) compara los resultados medios de las unidades justo por debajo y por encima del umbral.
- La diferencia da una estimación de los **efectos medios locales** del tratamiento del programa para los que están en el **umbral**.

Ejemplo de RDD: Prestaciones por desempleo (1)

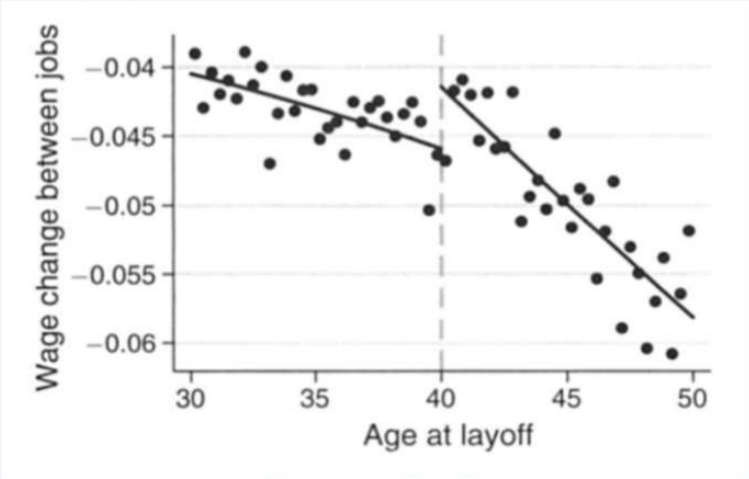
Nekoei/Weber (2017): ¿La ampliación de las prestaciones por desempleo mejora la calidad del empleo?

- **Interrogante investigado:** ¿Impacto de la prolongación de las prestaciones por desempleo en los salarios posteriores al desempleo?
- **Tratamiento:** El 1 de agosto de 1989, el gobierno austriaco aumentó la duración posible de las prestaciones por desempleo de 30 a 39 semanas para las personas de 40 años o más en el momento de la finalización del empleo.
- **Datos:** Datos administrativos sobre salarios y empleo para una muestra de $N \approx 1,7$ millones de despidos involuntarios.
- **Origen del sesgo de selección:** Los desempleados no afectados pueden tener menos experiencia en el mercado laboral y diferentes niveles de educación.
- **Método:** Comparar los salarios de reempleo de las personas algo más jóvenes y los mayores de 40 años cuando pierden su empleo.

Ejemplo de RDD: Prestaciones por desempleo (2)

– Gráficamente, el efecto del tratamiento es igual al salto vertical en la variable de resultado en el umbral.

Nekoei/Weber (2017): ¿La ampliación de las prestaciones por desempleo mejora la calidad del empleo?



Source: Fig. 3

– **Interpretación:** La ampliación del derecho a la prestación por desempleo aumenta los salarios de reemplazo en aproximadamente un 0,5 % para las personas que se encuentran en el umbral.

Conclusiones

- Para garantizar una **evaluación satisfactoria del programa...**
 - ...**implicar al evaluador** lo antes posible.
 - ...**planificar el método de evaluación** desde la concepción del proyecto.
 - **Importante:** Unos datos mejores ayudan mucho.
 - **Combinar diferentes fuentes de datos** (p. ej., datos administrativos y de estudios) puede ser útil en muchas situaciones (pero también puede llevar mucho tiempo).
- Puede ser útil pensar en **recoger nuevos datos** específicamente para la evaluación del programa.

«Puede hacerlo todo. Pero no se puede hacer todo y, desde luego, no se puede hacer todo al mismo tiempo.

1. Introducción	2. Marco de evaluación	3.1. Experimentos	3.2. Adecuación	3.3. DID	3.4. RDD	4. Conclusiones	Apéndice
○○○○	○○○	○○	○○○	○○○○○	○○○	○●	○○○○○○○○○○

Gracias por su atención