

El Efecto PISA: más allá de la Metodología y la Comparación

Evaluaciones de Impacto e Inferencia Causal sobre Bases de Datos Internacionales

Daniel Santín González
Profesor Titular de Universidad
Departamento de Economía Aplicada VI
Universidad Complutense de Madrid
dsantin@ccee.ucm.es

Objetivos

- Introducción a la evaluación de impacto
- Principales técnicas de inferencia causal
- Aplicaciones en bases de datos internacionales
- Principales resultados e ideas

Algunas premisas (I/II)

- Las Políticas Educativas deberían estar basadas en evidencias científicas que demuestren una relación causal entre la política y el indicador perseguido (BM, BID, ONU, OCDE...).
- Las Políticas Educativas no deberían estar basadas en creencias, buena fe, ideologías o en copiar otras experiencias de dudosa efectividad.
- Distintas preferencias sociales en el concepto de igualdad de oportunidades pero práctica unanimidad en el concepto de eficiencia paretiana.
- Las técnicas de inferencia causal permiten establecer una relación causal y **cuantificar el impacto** que una política pública (educativa) tiene sobre una población.

Algunas premisas (II/II)

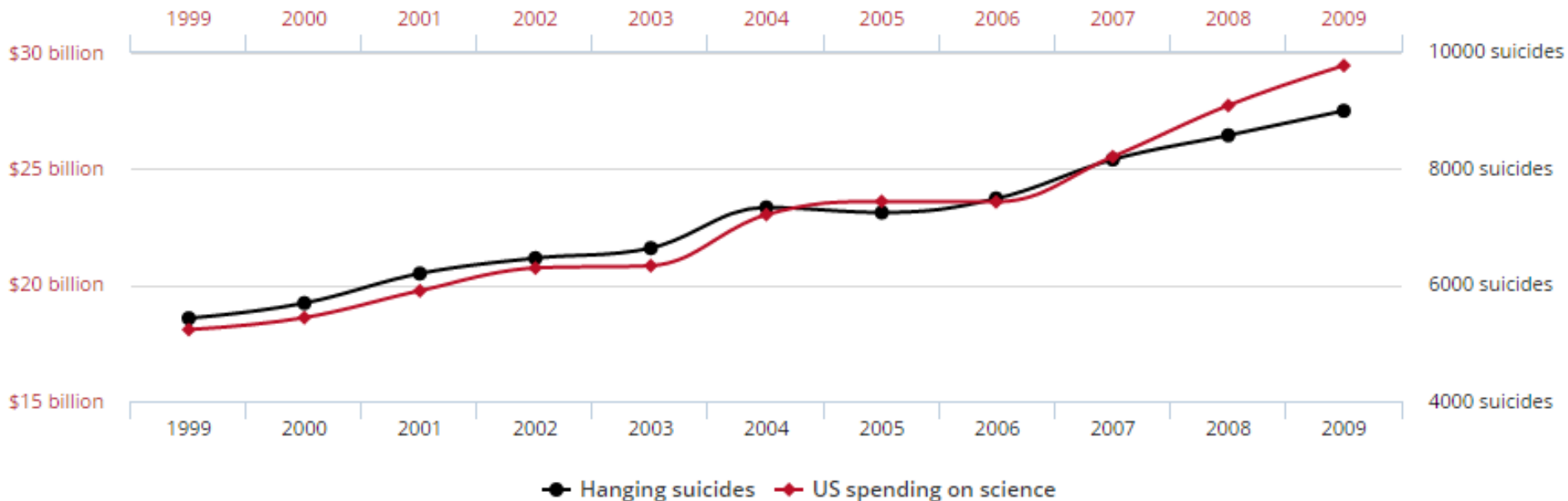
- Correlación no es causalidad.
- Obtener evidencia científica de causalidad del tipo $X \rightarrow Y$ es muy costoso (tiempo, recursos, análisis, etc.) pero también muy valioso para tomar decisiones.
- Antes de comenzar...plantearse si es más costoso evaluar o no evaluar.
- De hecho, en bastantes ocasiones sólo será posible hacer estudios correlacionales basados en regresiones múltiples (una buena práctica debería ser avisar al lector).

<http://www.tylervigen.com/spurious-correlations>

Correlación y Causalidad

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation

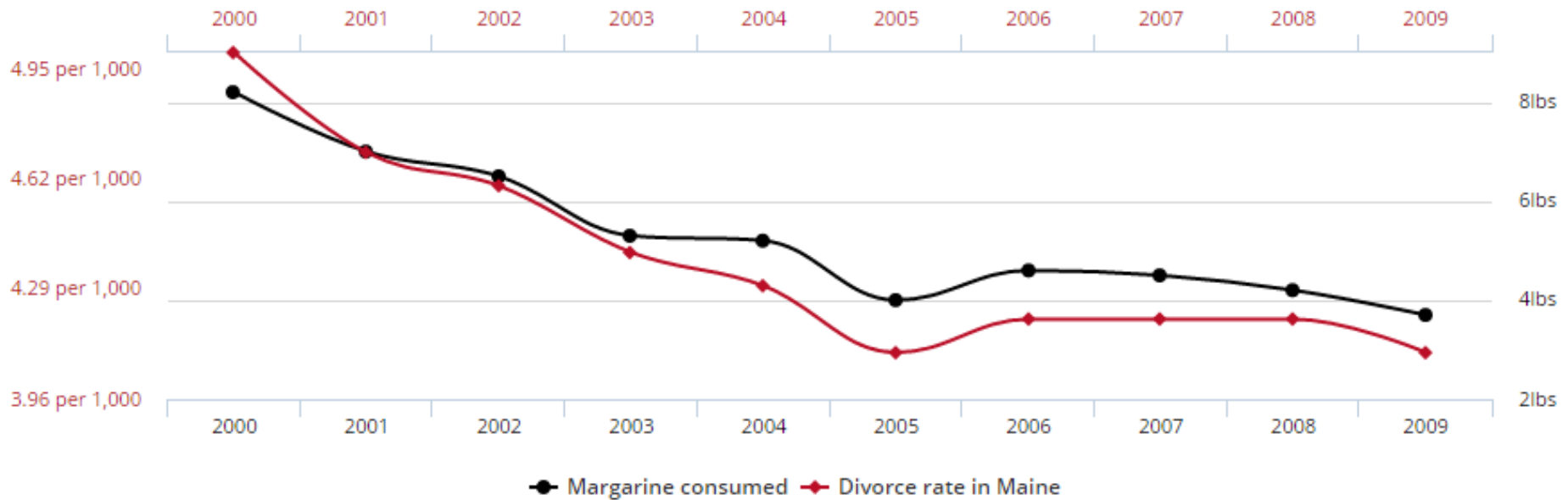
Correlation: 99.79% ($r=0.99789126$)



Correlación y Causalidad

Divorce rate in Maine correlates with Per capita consumption of margarine

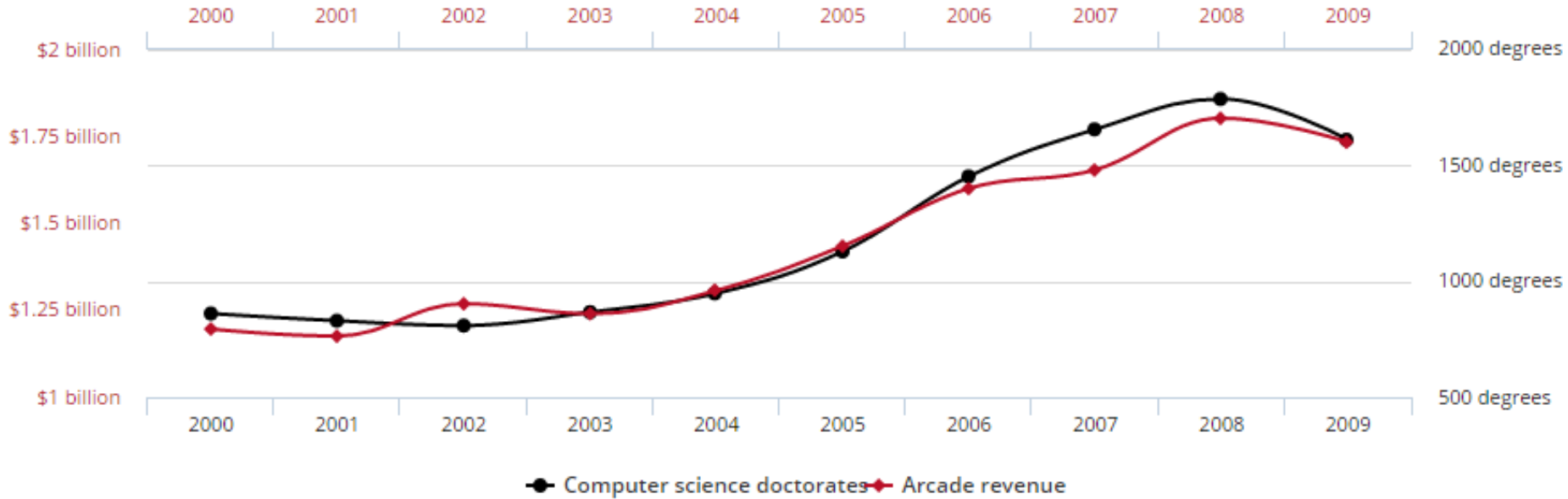
Correlation: 99.26% ($r=0.992558$)



Correlación y Causalidad

Total revenue generated by arcades correlates with Computer science doctorates awarded in the US

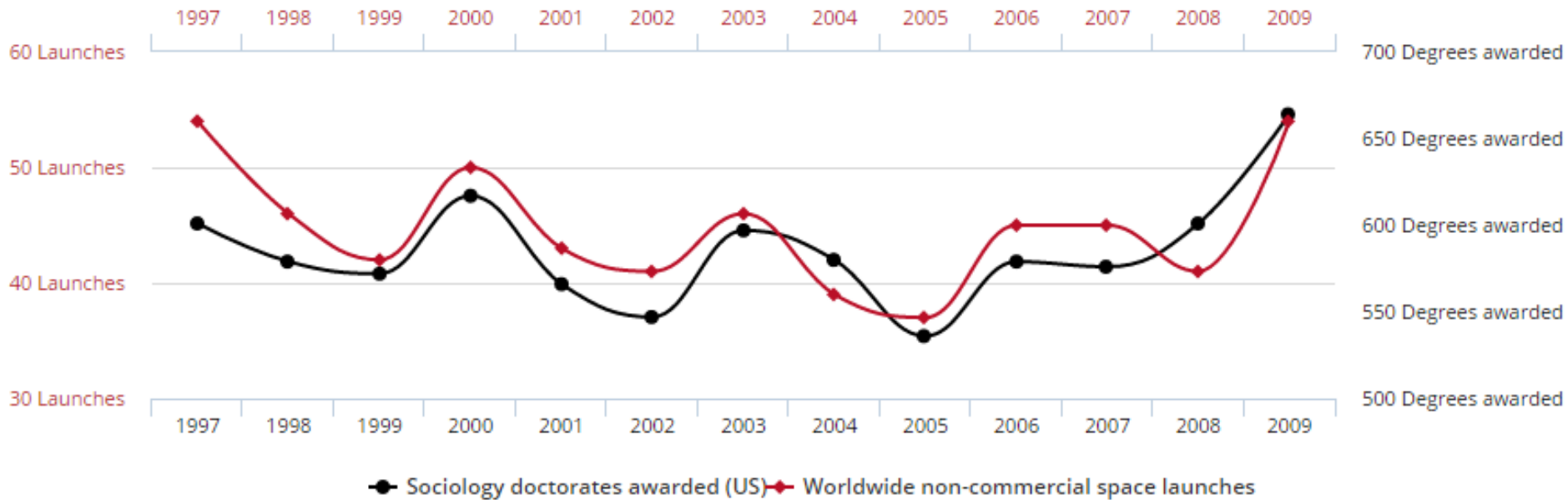
Correlation: 98.51% ($r=0.985065$)



Correlación y Causalidad

Worldwide non-commercial space launches correlates with Sociology doctorates awarded (US)

Correlation: 78.92% (r=0.78915)



Inferencia causal



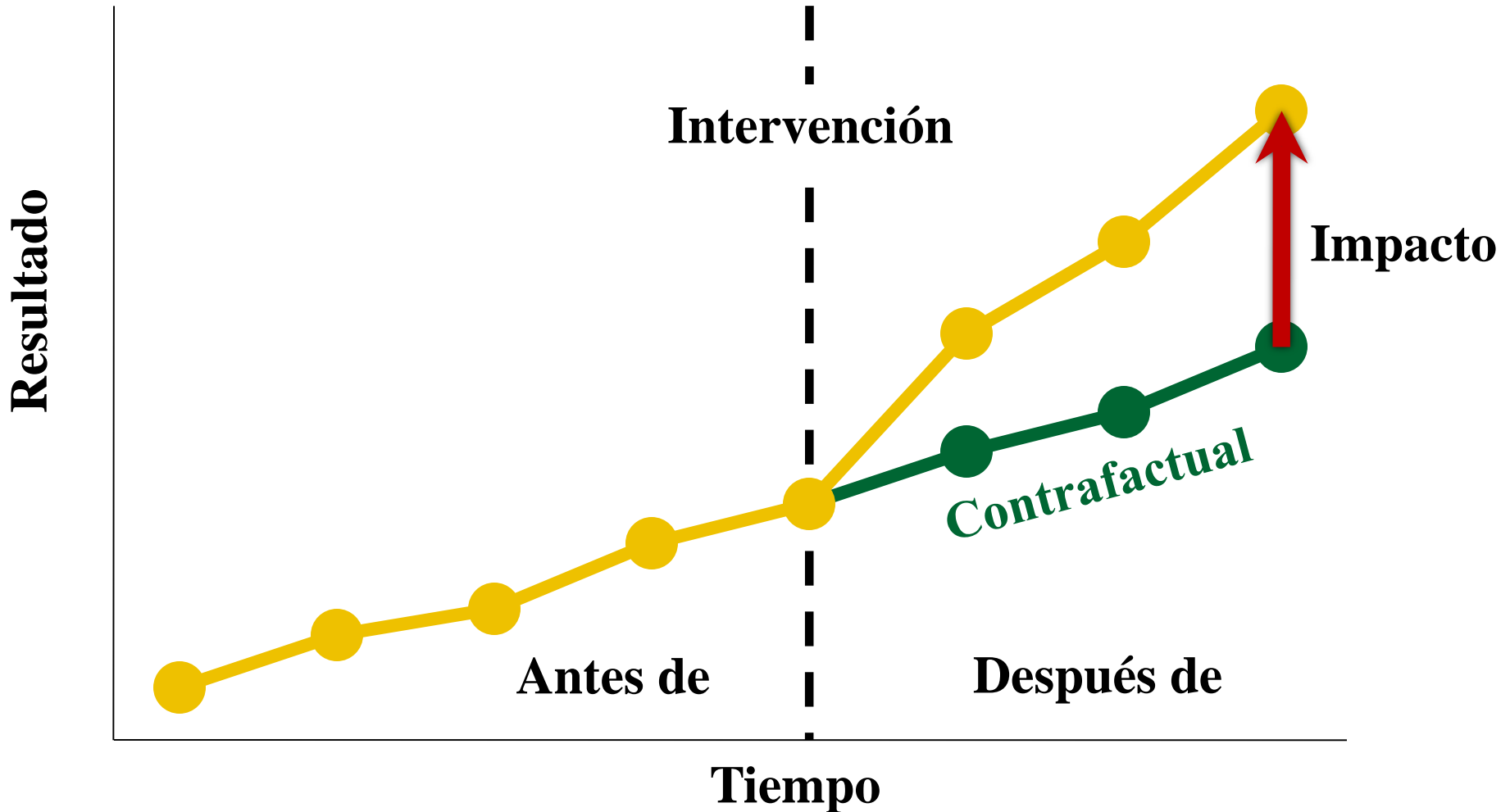
Situación contrafactual

- Necesitamos definir un grupo **contrafactual** que sea **esencialmente igual** que el grupo de participantes menos en una cosa: su exposición al programa que se está evaluando.

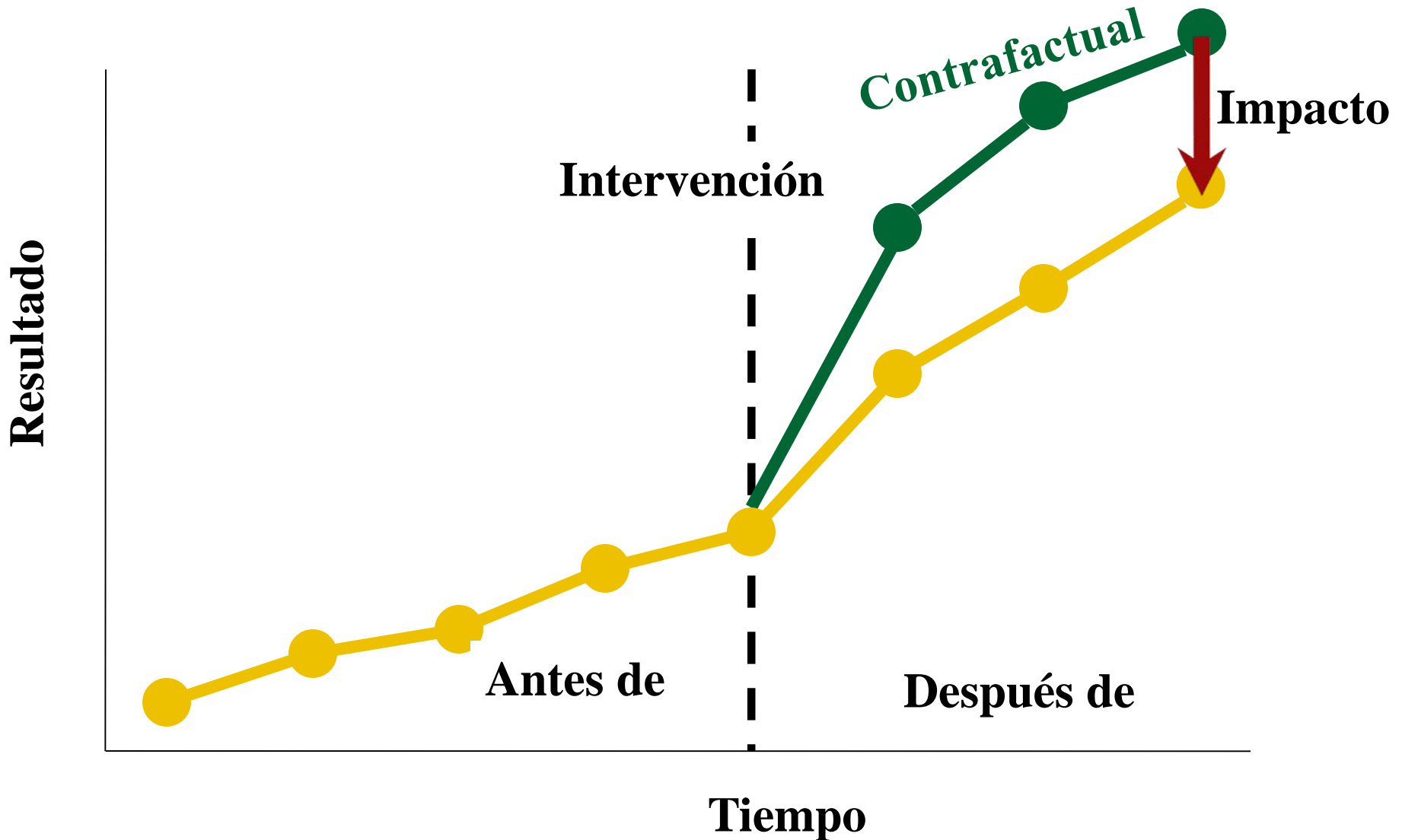


Source: Andrew Fraker

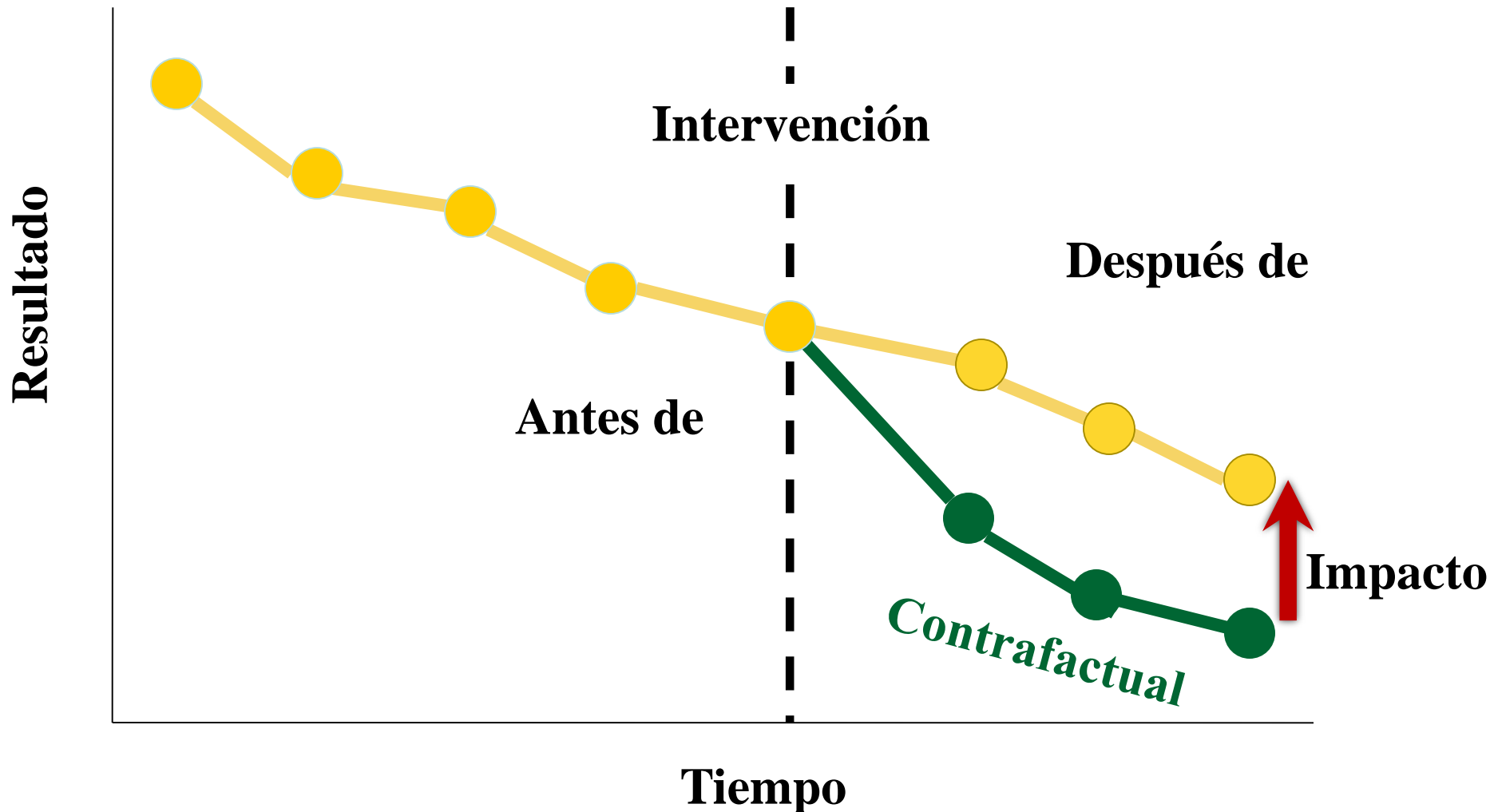
Impacto - escenario #1 (positivo)



Impacto - escenario #2 (negativo)



Impacto - escenario #3 (positivo)



Diseños de Evaluación

- Para llevar a cabo una evaluación de impacto debemos encontrar una **estrategia de identificación** mediante alguna de las siguientes técnicas:
 - Experimentos sociales controlados, ensayos aleatorizados y loterías con sobre demanda (escasos)
 - Diseños cuasi-experimentales (principales)
 - Regresiones en discontinuidad
 - Diferencias en diferencias
 - Variables instrumentales
 - Propensity Score Matching
 - ...Bases de datos internacionales

De correlación a causalidad (I/II)

- Alta prioridad por trabajos causales
- Sustitución progresiva de trabajos tradicionales (análisis multivariante, OLS) por econometría más sofisticada.
- ¿Qué podemos hacer con las bases de datos internacionales?
- Repaso de la literatura
 - Estrategias de identificación
 - Problemas analizados
 - Resultados obtenidos
 - Necesidades de información

CAUSAL INFERENCE ON EDUCATION POLICIES: A SURVEY OF EMPIRICAL STUDIES USING PISA, TIMSS AND PIRLS

José M. Cordero*  and Víctor Cristóbal

University of Extremadura

Daniel Santín

Complutense University of Madrid

Estrategia de búsqueda (I/II)

- ERIC (Educational Resources Information Center)
- ISI Web of Science
- Econlit (American Economic Association)
- ABI/Inform Global
- Elsevier Scopus
- Google Scholar.
- Papers en inglés publicados entre 2004 y 2016
- Deben usar PISA, TIMSS o PIRLS...el interés fundamental es qué podemos hacer en economía con estas bases de datos.

Estrategia de búsqueda (II/II)

- Palabras clave en el abstract:
 - PISA, TIMSS, PIRLS, large-scale assessment,
 - Causal inference, identification strategy, exogenous variation,
 - Instrumental variables, propensity score matching, difference-in-difference, fixed effects,
 - Cross-country, comparative study, student performance and achievement.
- 180 papers...pero después de una criba **66 trabajos** que clasificamos por la estrategia de identificación y por el tema estudiado.

Análisis de bases de datos internacionales

PISA

PIRLS

TIMSS

Variable 1

Variable 2

Variable 3

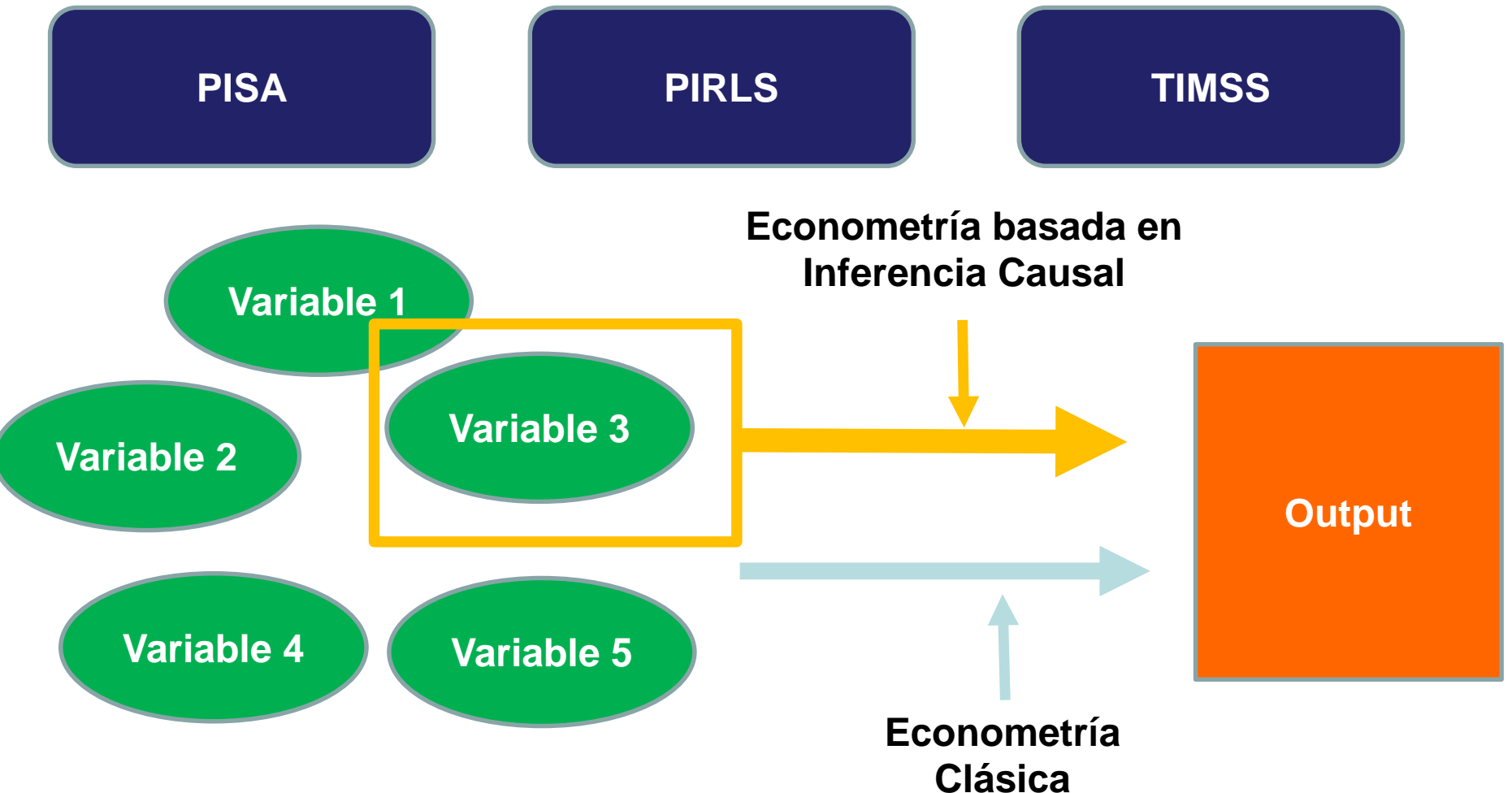
Variable 4

Variable 5

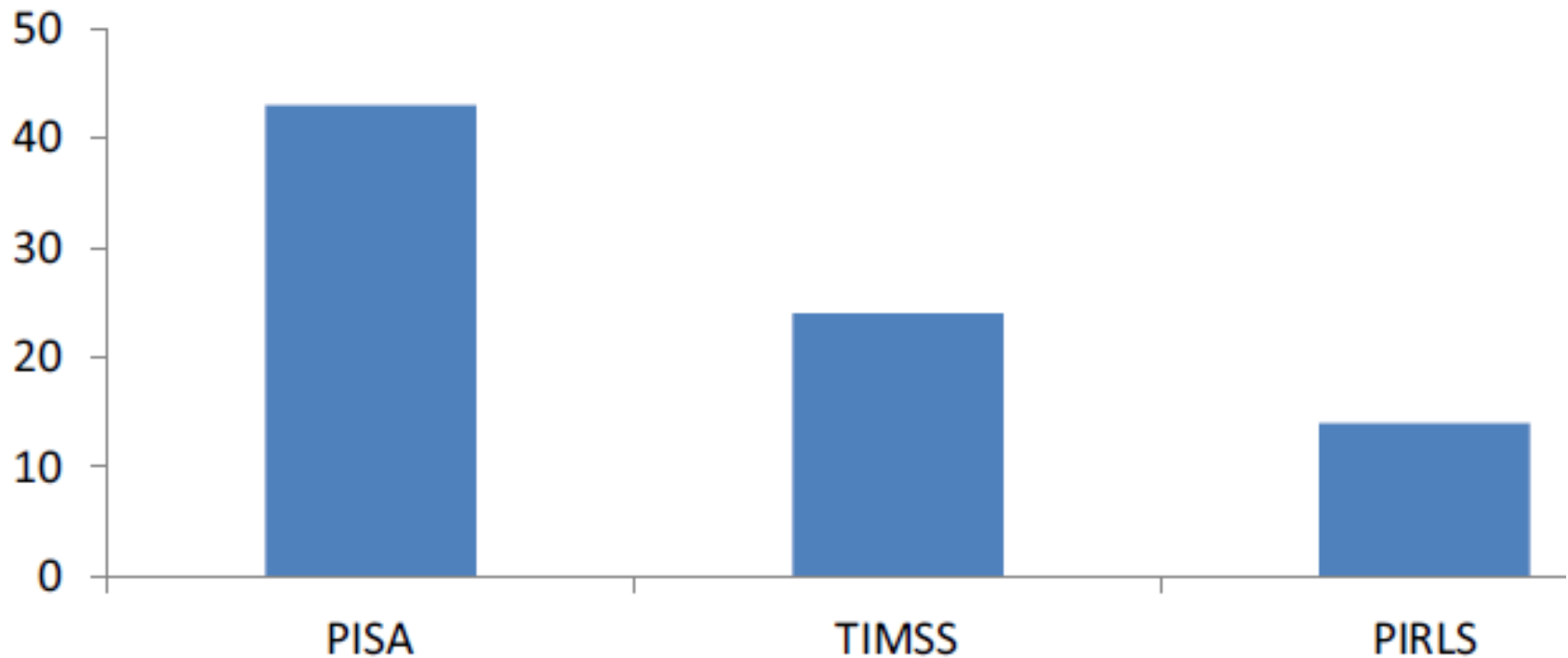
Econometría basada en
Inferencia Causal

Output

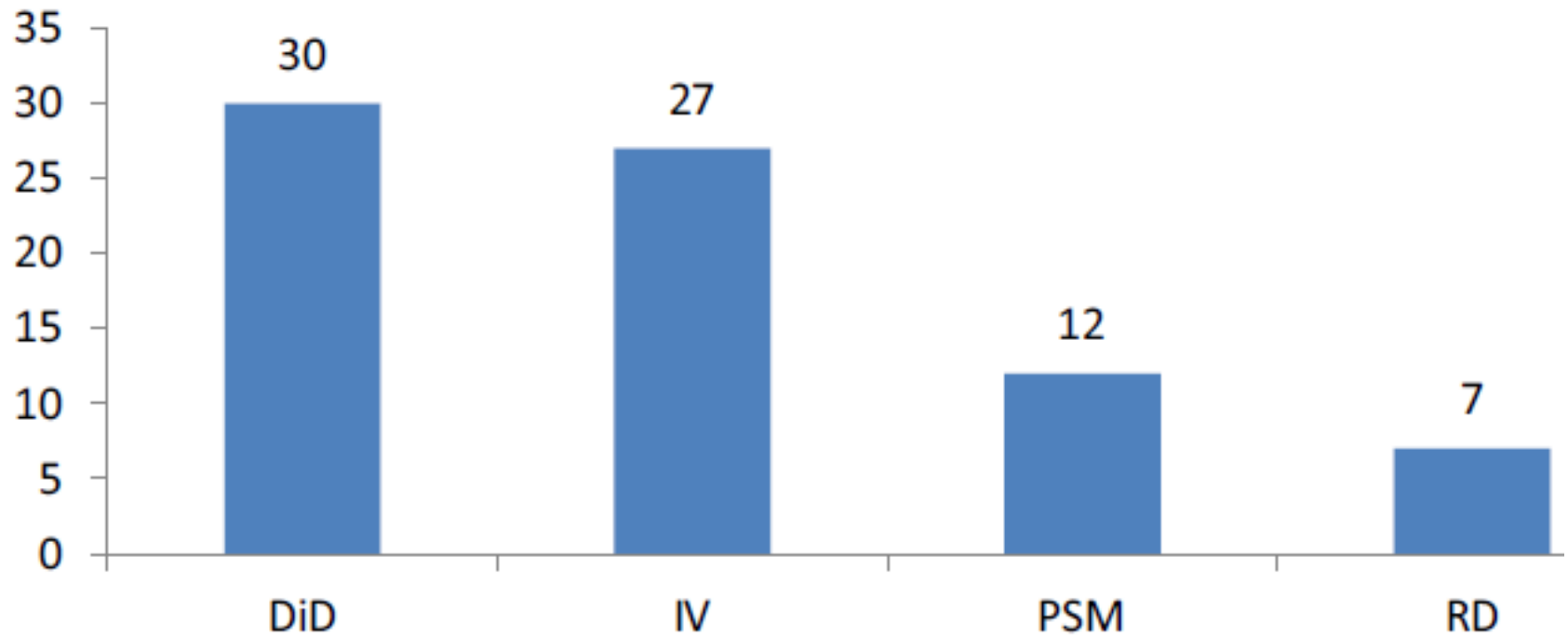
Econometría
Clásica



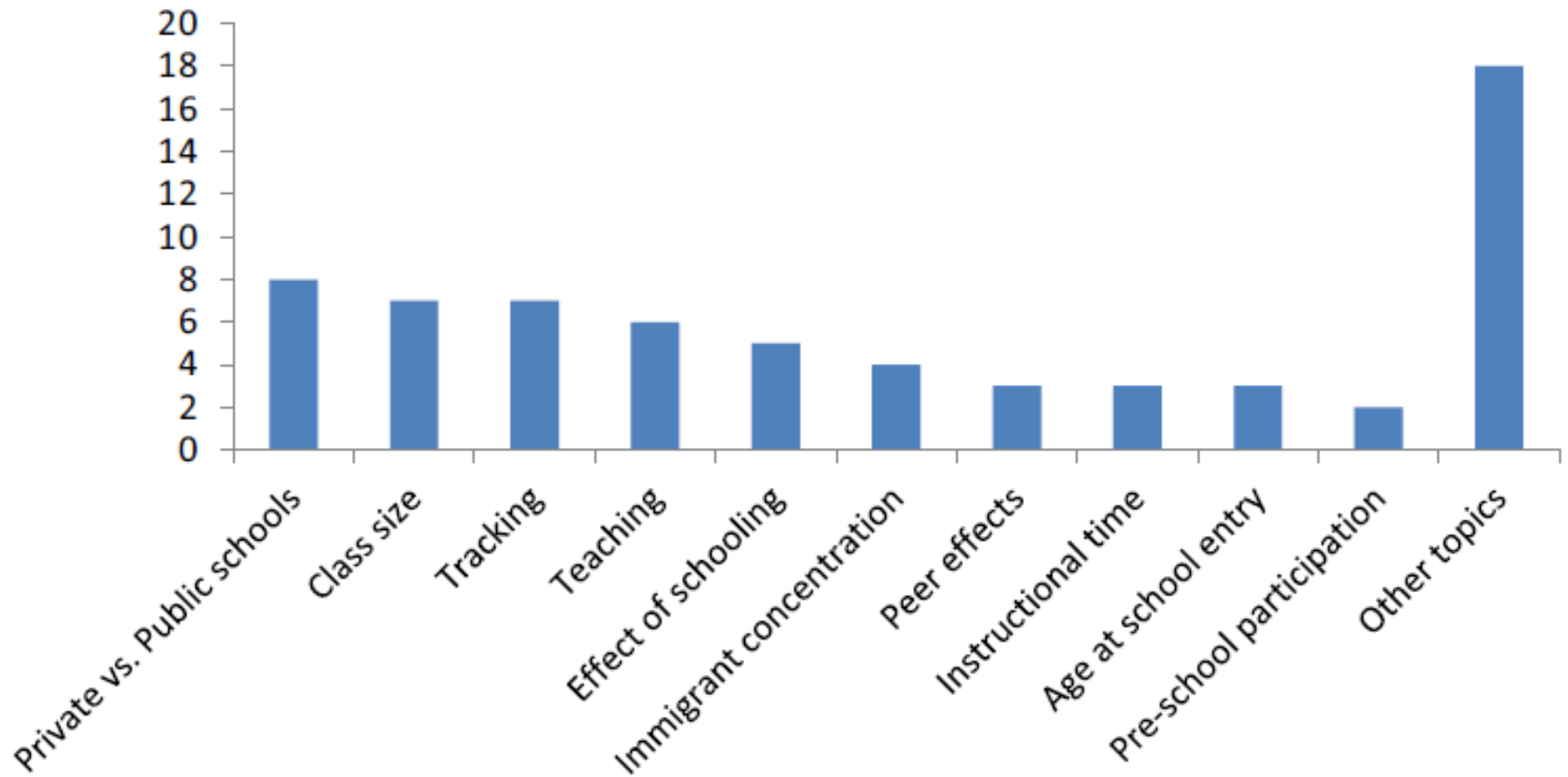
Bases de datos



Método de Estimación



Variable relevante



VARIABLES INSTRUMENTALES

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon$$

- **IDEA:** La variable relevante (D) no es exógena.
- **CLAVE:** Busca una variable Z correlacionada con D pero no correlacionada con Y .
- Ejemplos (PISA, TIMSS, PIRLS):
 - Entrada al sistema educativo → trimestre de nacimiento.
 - Asistencia a concertada → tamaño de la población; porcentaje de población católica en 1900.
 - Tamaño de la clase → Tamaño medio de la clase en distintos cursos
 - Competencia escolar (número de distritos escolares) → número de ríos alrededor.
 - Se creativo!!
 - Ver Barrio Sésamo en los 60s → Distancia del distrito a la antena de UHF que emitía el programa.

Variables Instrumentales

Approach	Description	Strengths	Weaknesses
Instrumental variables (IV)	Sometimes nature or the legal framework leads to exogenous sources of variation correlated with the treatment but uncorrelated with the dependent variable.	The method exploits a partial random assignment that reproduces a natural experiment. It provides even more robust results than other methodological approaches.	It is mostly quite difficult to find a good endogeneity-free instrument from international databases.

Regresión en discontinuidad

- Desde un punto de vista econométrico simple (Sharp design):

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 \text{Indice} + \beta_3 X + \varepsilon$$

- donde “índice” es el valor entre 1 a 100 que presenta esa familia (*running variable*) y:

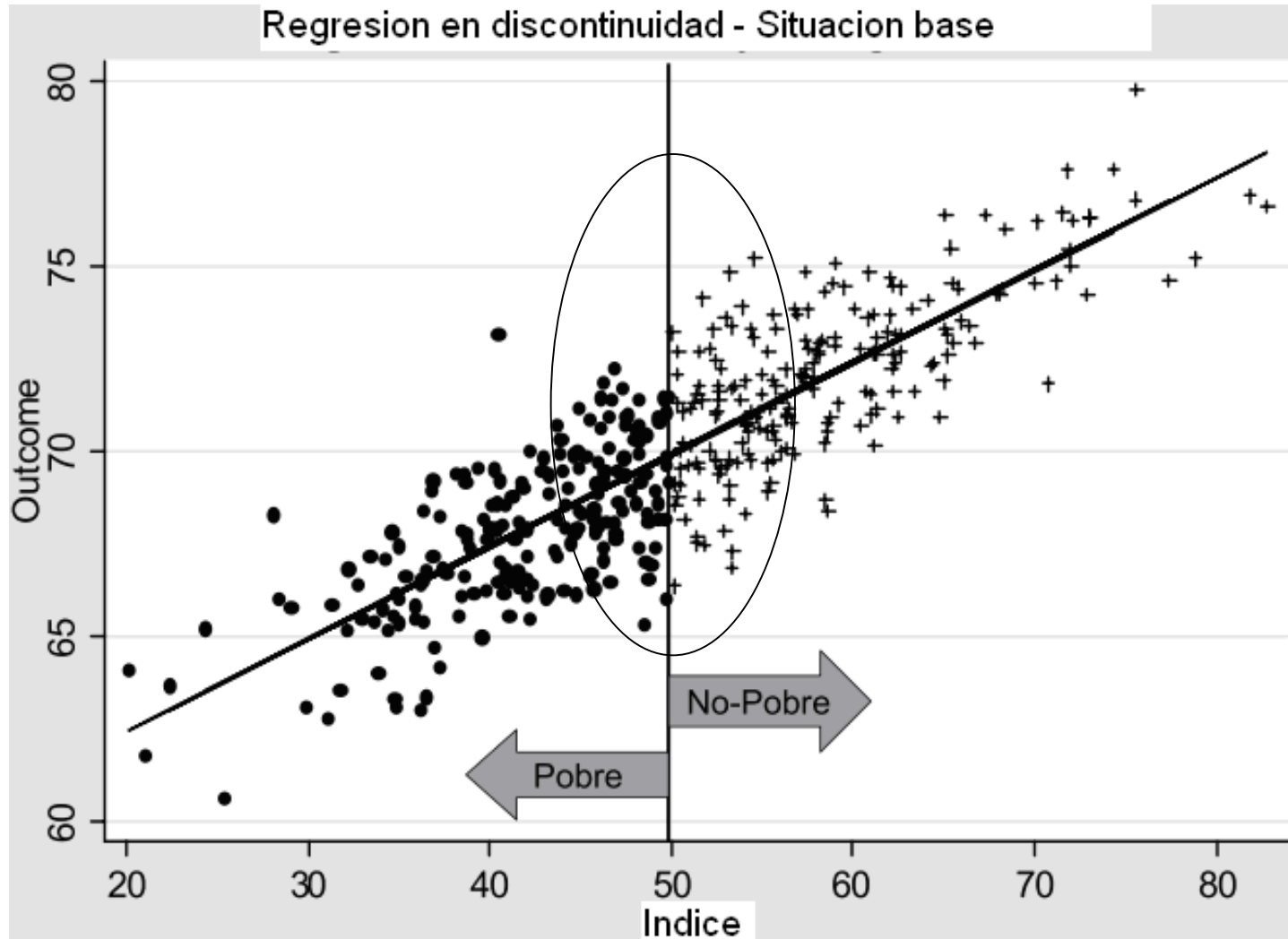
$$D = \begin{cases} 1 & \text{si recibe el tratamiento} \\ 0 & \text{si no recibe el tratamiento} \end{cases}$$

- La regla para determinar el tratamiento ($D=1$) basada en punto de corte:

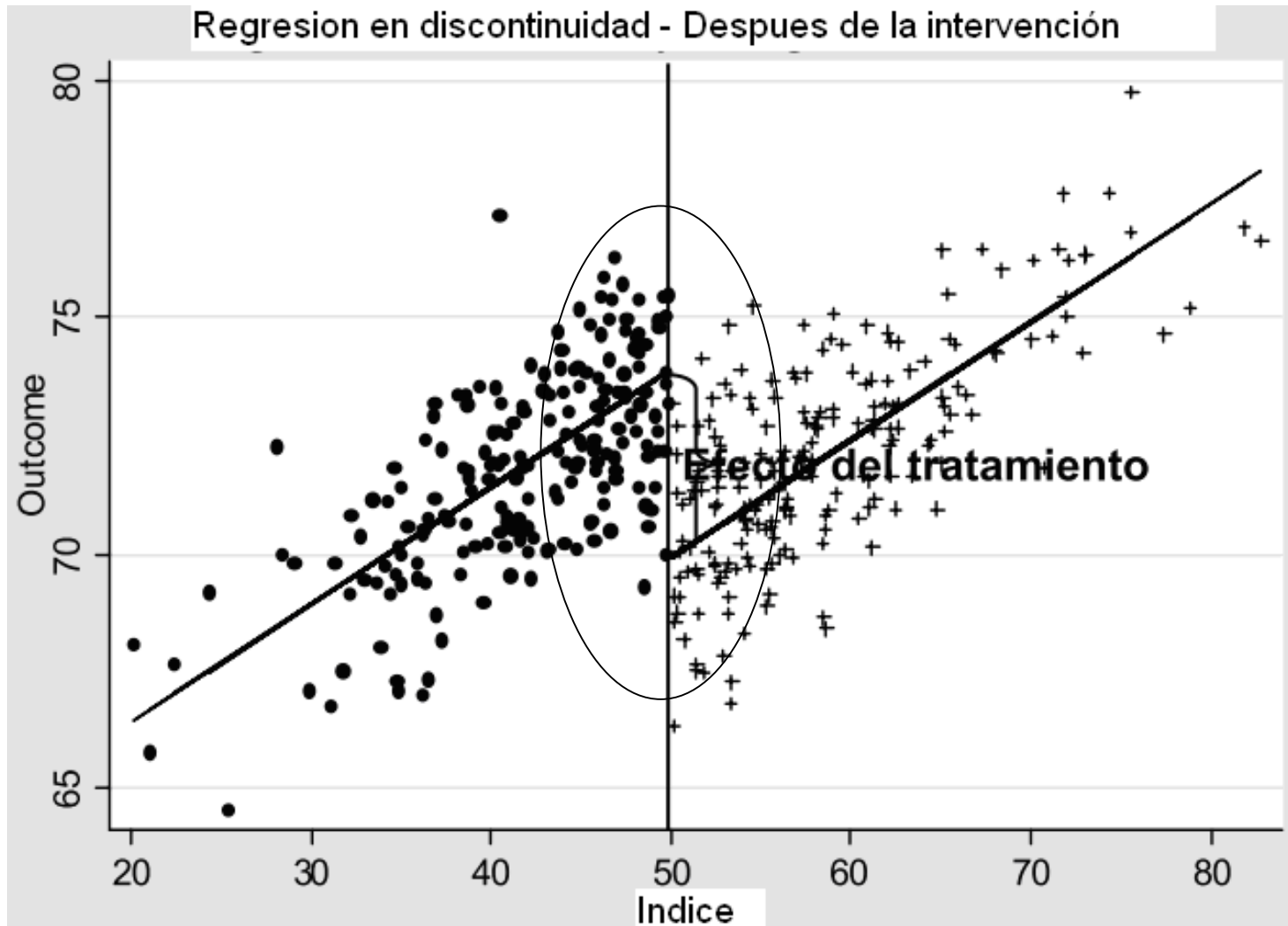
$$D_i = 1 \Leftrightarrow \text{Indice} \leq 50$$

$$D_i = 0 \Leftrightarrow \text{Indice} > 50$$

Regresión en Discontinuidad



Regresión en Discontinuidad



Regresión en Discontinuidad

- **IDEA:** La variable relevante (D) presenta un punto de corte
- **CLAVE:** Busca una política o regulación basada en el punto de corte. En muchos casos la regla es difusa y necesita combinación con IV.
- Ejemplos (PISA, TIMSS, PIRLS):
 - Tamaño de clase → Países que aplican reglas de Maimonides.
 - Entrada al sistema educativo → Año natural. En TIMSS aprovechan la existencia de dos cursos adyacentes.
- Es el método menos veces aplicado usando bases de datos internacionales.
- Normalmente los puntos de corte son difusos por lo que el uso de la técnica se realiza a través de IV.

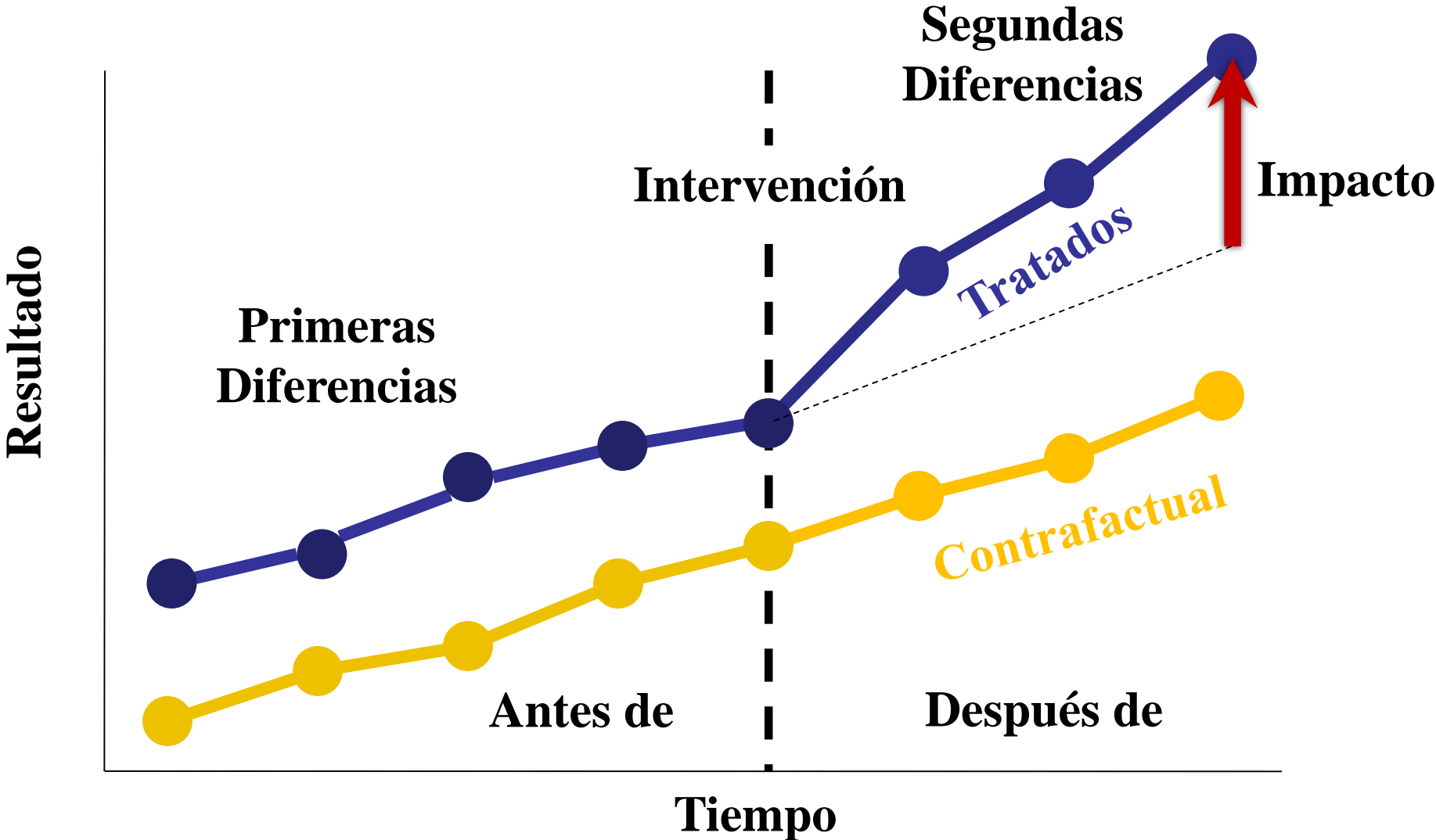
Regresión en Discontinuidad

Approach	Description	Strengths	Weaknesses
Regression discontinuity designs (RDD)	Participation is decided by an exogenous cut-off point, normally defined by an education law requirement.	The cut-off point reproduces a random experiment. It is easy to apply and provides robust results. It works well with educational policies based on rules, such as grants, entry criteria etc.	Results are average local treatment effects in the sense that they could not be generalized for individuals that are far from the cut-off point. Most of times we find a fuzzy RDD.

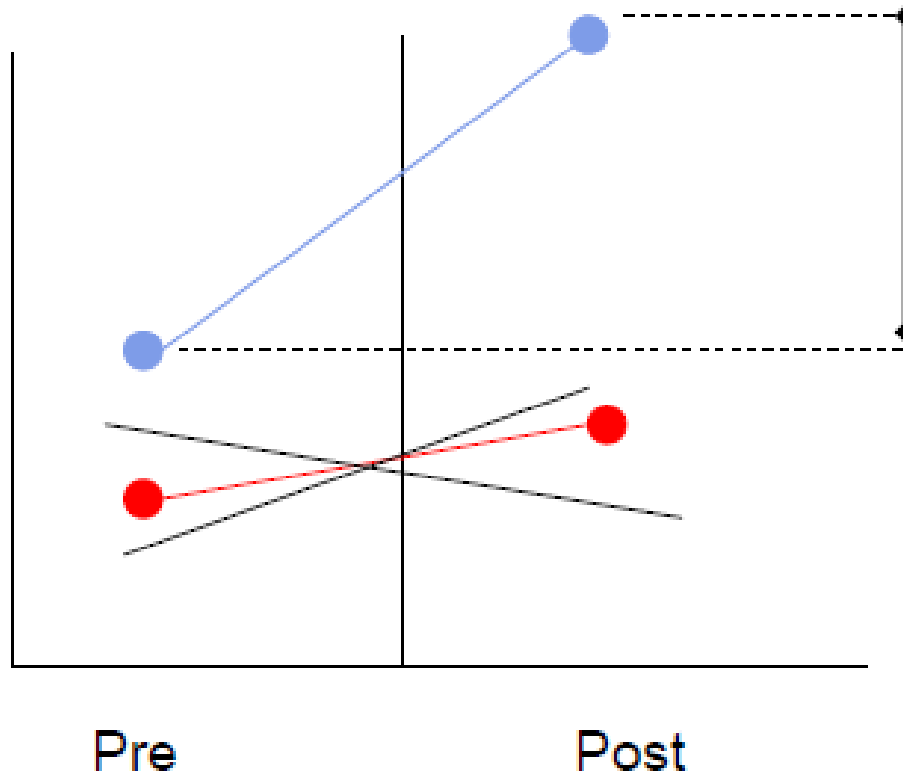
Diferencias en Diferencias

- **IDEA:** Dos grupos son diferentes pero de forma exógena uno es sometido a un tratamiento y el otro no.
- Tenemos información “antes del tratamiento” y “después del tratamiento” para los dos grupos
- **CLAVE:** Busca programas implementados en algunas regiones pero no en otras.
- En PISA cada oleada es un pseudo-panel por lo que el análisis se suele agregar a nivel regional.

Diferencias en Diferencias

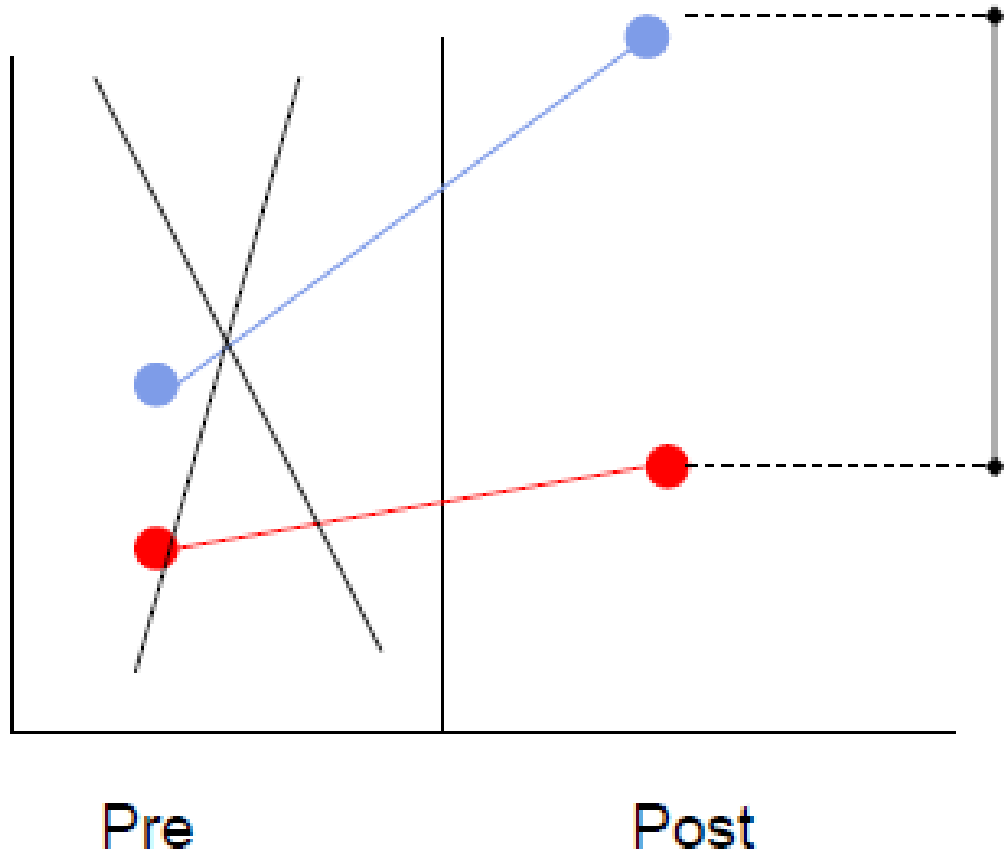


Diferencias en Diferencias: Falacia 1



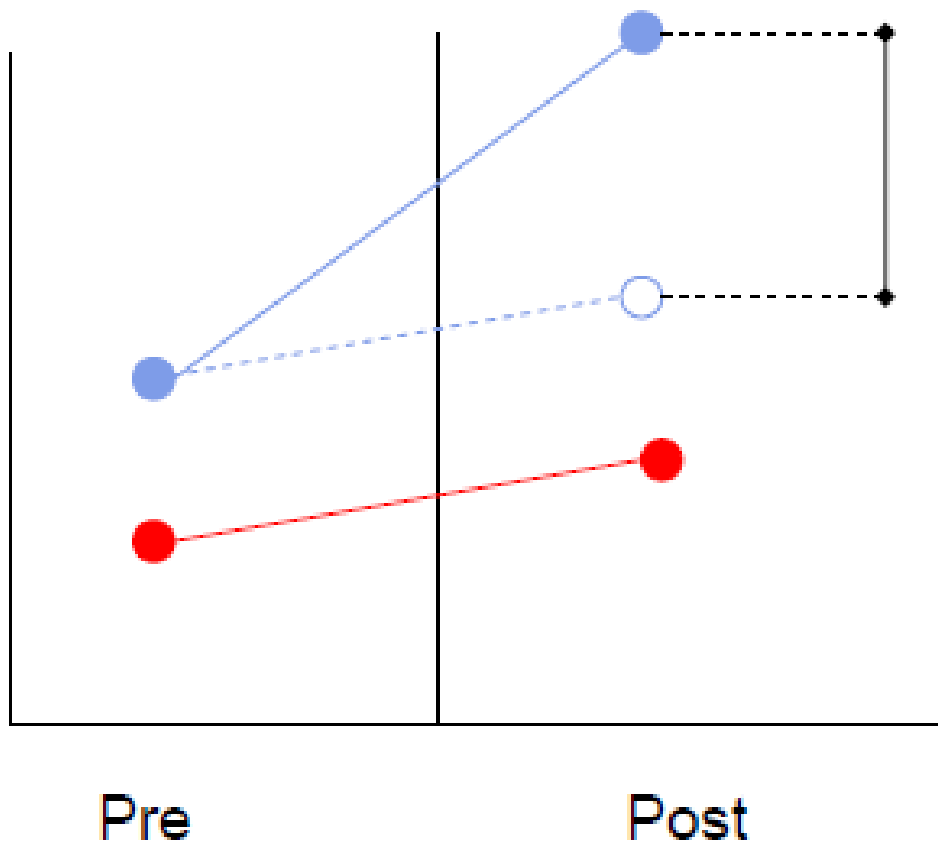
Effect of program using only pre- & post- data from T group (ignoring general time trend).

Diferencias en Diferencias: Falacia 2



Effect of program using only T & C comparison from post-intervention (ignoring pre-existing differences between T & C groups).

Diferencias en Diferencias: efecto causal



Effect of program
difference-in-difference
(taking into account pre-
existing differences
between T & C and
general time trend).

Diferencias en Diferencias

- **Ejemplos:**
 - Tracking: Los países no tienen tracking en primaria pero algunos países si tienen en secundaria y otros no.
 - Asistencia a infantil: Comparación a los 15 años de los resultados de alumnos antes y después de una reforma. Felfe et al. (2015) explotan en PISA la implantación de la LOGSE a distintos ritmos en las CCAA encontrando un impacto significativo de más años de educación infantil.
 - Pruebas de rendimiento: Un grupo de regiones no tiene pruebas de rendimiento y algunas deciden introducirlas.
- Ha sido el método más veces aplicado

Regresión en Discontinuidad

Approach	Description	Strengths	Weaknesses
Differences in differences (DiD)	'Before' and 'after' information is required for the treated and the counterfactual groups. The treatment should be exogenous for the treated group.	Once the information is available and the equal trends assumption is verified before applying the treatment, the method is easy to apply and provides robust results.	Data demanding in terms of 'pre' and 'post' periods. It is crucial to demonstrate the equal trends assumption. For international databases, this probably requires the linkage of different waves.

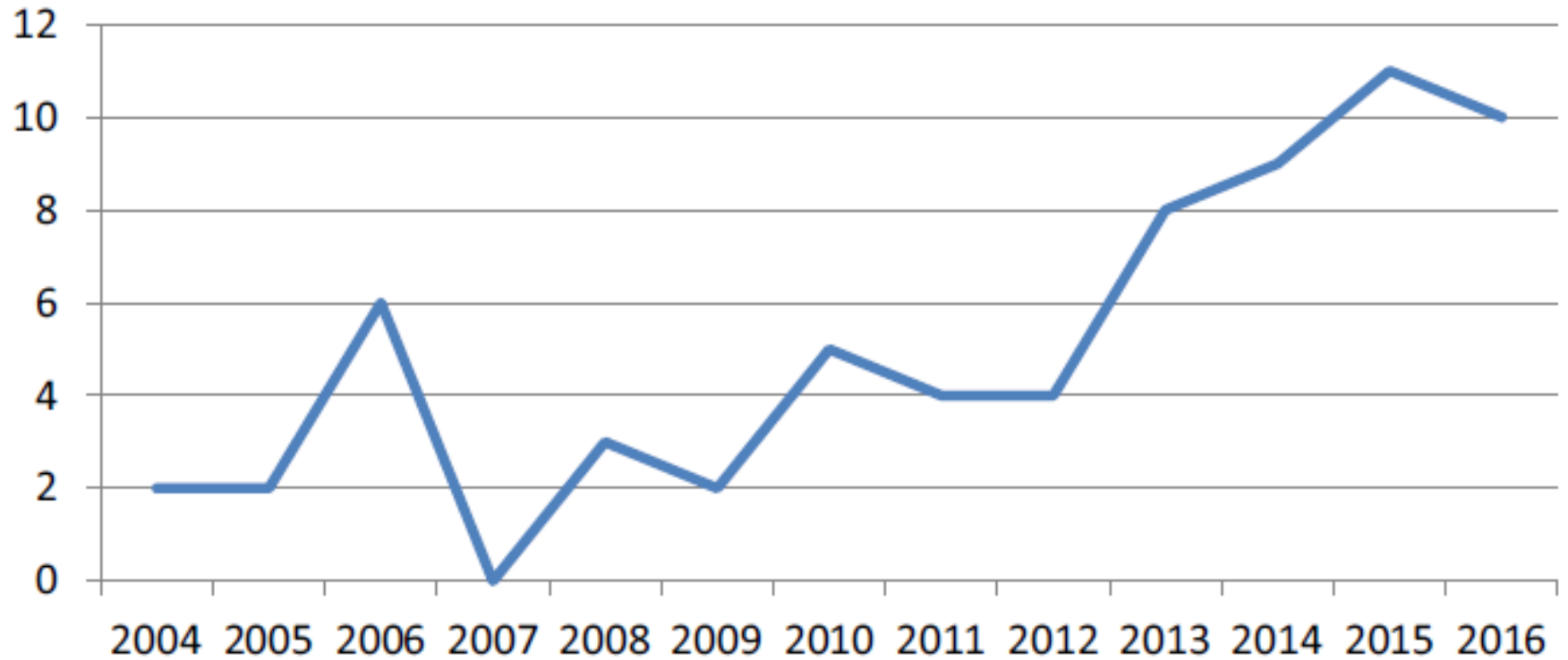
Propensity Score Matching

- En desuso...
- IDEA: Emparejar unidades tratadas y no tratadas por aquellas características observables que decidieron la probabilidad de recibir el tratamiento. Método no experimental...no empareja por inobservables.
- CLAVE: Si no puedes aplicar ninguno de los métodos anteriores aplica PSM para reducir el sesgo.
- **Ejemplos:** (emparejados por probabilidad de tratamiento)
 - Tracking: Alumnos que lo siguen vs no
 - Pública - concertada: alumnos que siguen cada sistema.
 - Infantil: Más años Vs menos años

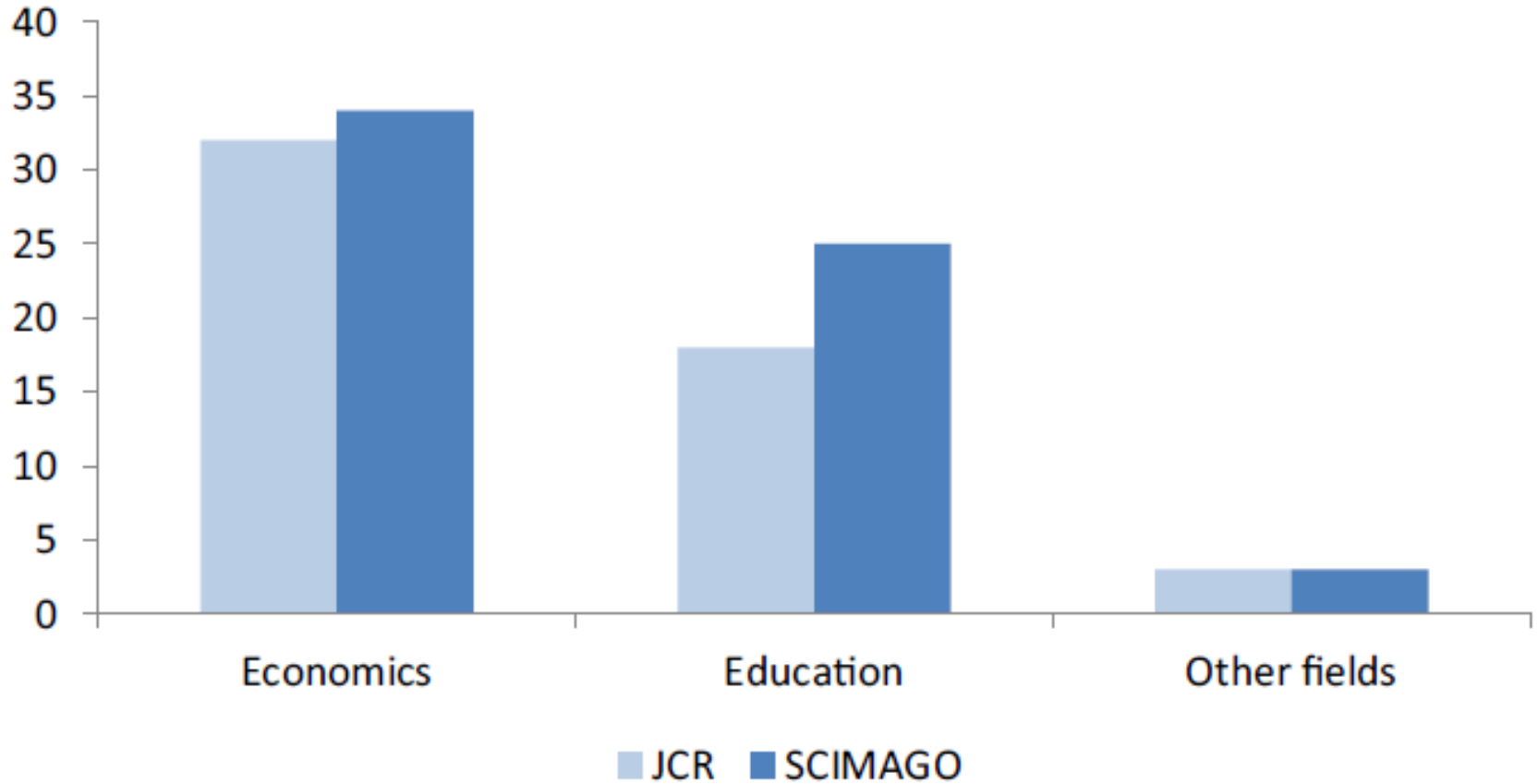
Propensity Score Matching

Approach	Description	Strengths	Weaknesses
Propensity score matching (PSM)	<p>Beneficiaries are matched with control individuals using prior-to-treatment observed covariates. This requires an estimation of the probability of belonging to the treated group for all individuals. Then, the estimated probabilities are used to match pairs of treated individuals and control individuals that have a similar probability of being treated but are in the control group.</p>	<p>PSM improves causal estimations with respect to using all untreated individuals as a control as long as unobservable variables correlate with observables. Whenever this assumption holds and treated and control individuals have the same distribution on unobservable variables, PSM mitigates the endogeneity problem.</p>	<p>PSM is a non-experimental approach because there is no randomization in the treatment assignment. It is mostly unreliable to assume that the unobservable variables of students or parents affecting both the treatment and the results will be equally distributed in the treated and untreated groups.</p>

Tendencia

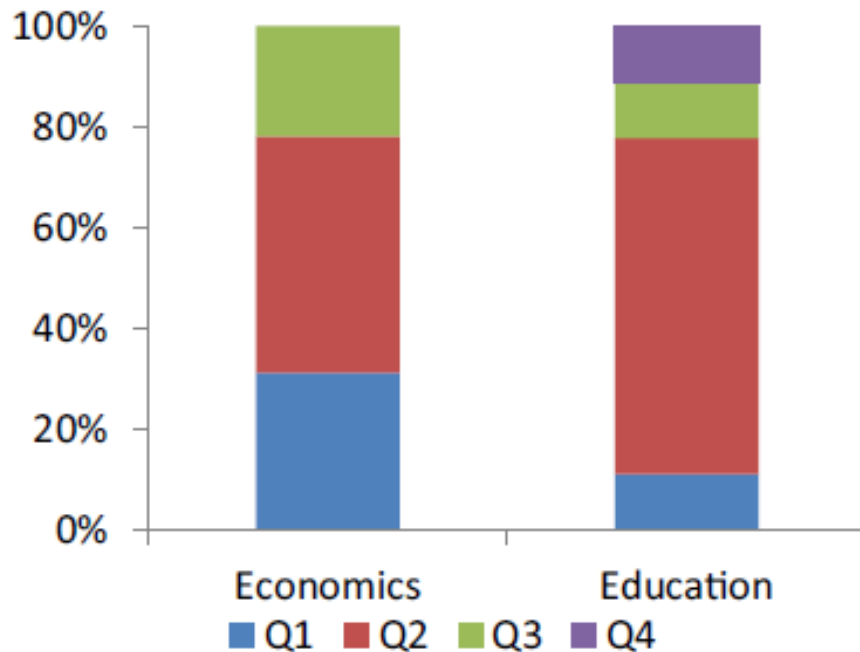


Revistas científicas

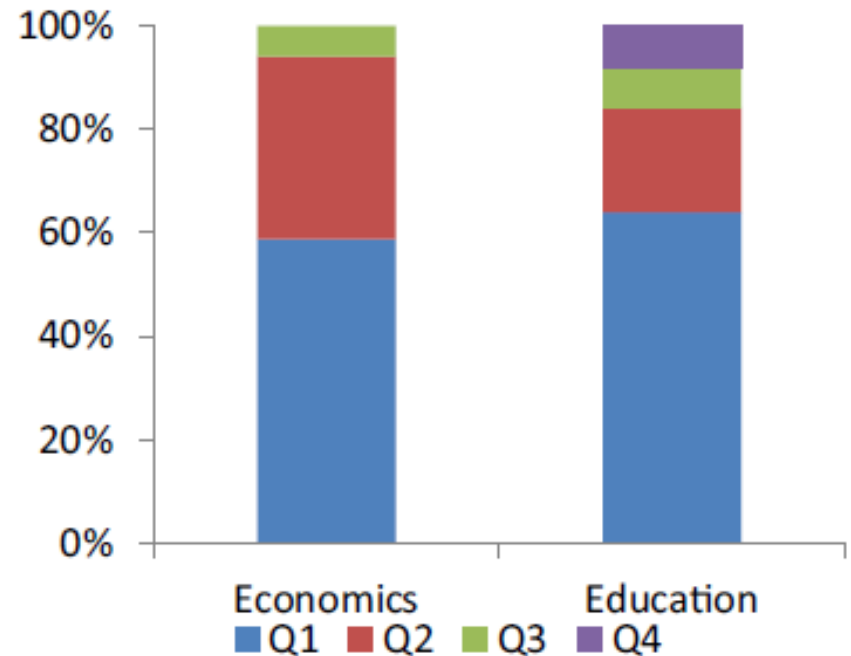


Impacto de las Revistas

(a) JCR



(b) SCIMAGO



Conclusiones (I/II)

- En ocasiones (con mucha creatividad) PISA puede ser usada para evitar el problema que crea la endogeneidad.
- DiD parece el método más fácil de aplicar ya que cada vez están disponibles más oleadas y más desagregadas.
- 66 publicaciones en 12 años con las tres bases de datos internacionales (5 por año) no permiten todavía encontrar ningún patrón claro de política educativa
- Los estudios que muestran correlaciones deberían decir explícitamente que no son causales.
- PISA NO resuelve la necesidad social de evaluación causal.

Conclusiones

- Necesitamos:
 - Bases de datos longitudinales
 - Saber cómo se asignan los alumnos a los grupos.
 - Varios grupos por centro
 - Voluntad política de evaluar
 - Que la sociedad civil demande y exija evaluaciones de calidad
 - Innovación docente, programas piloto
 - Obligar a que cada nuevo programa lleve acompañado el método por el que será evaluado.

Muchas Gracias!!

Daniel Santín González
Profesor Titular de Universidad
Departamento de Economía Aplicada VI
Universidad Complutense de Madrid
dsantin@ccee.ucm.es