Inferencia en Estudios Observacionales

2025-05-07

Intro

Intro

Hoy vamos a ver otros dos tópicos importantes relacionados con la inferencia en contextos no experimentales.

También conoceremos un debate que se dio en la literatura académica sobre métodos experimentales y no experimentales, empezando con Lalonde (1986).

Errores Agrupados II

Errores Agrupados II

En la clase pasada, discutimos el papel de errores agrupados en ensayos aleatorios.

En general, si tenemos un experimento a nivel de grupo, los errores tienden a ser mayores. Tenemos menos observaciones independientes de las que imaginamos.

Errores Agrupados II

Ese problema también ocurre en contextos no experimentales.

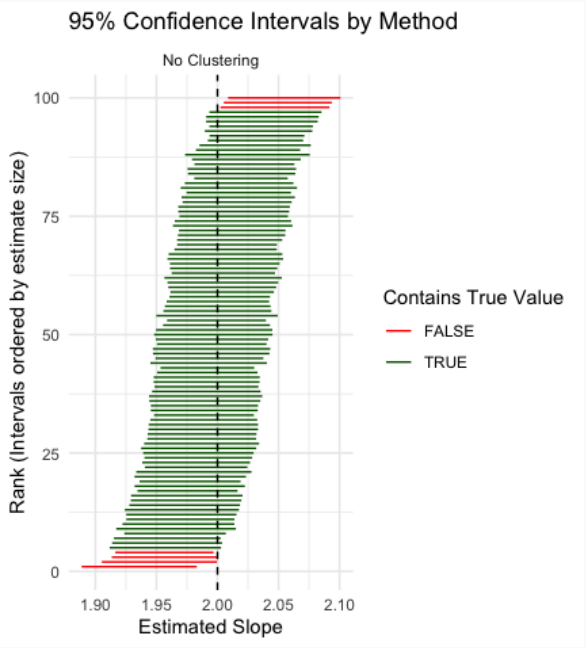
Es muy común que una muestra incluya una estructura jerárquica: diferentes personas en diferentes comunas, o regiones por ejemplo.

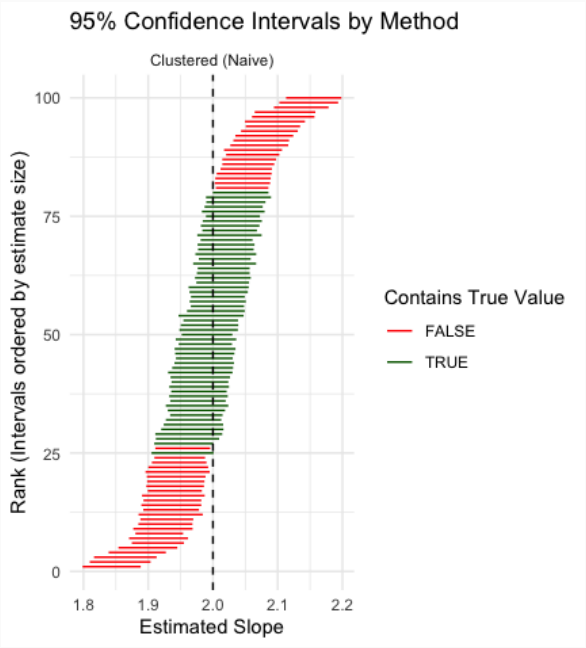
Si existen shocks o diferencias no observadas a un nivel más alto, tenemos un problema similar: los errores no son i.i.d., y por tanto el estimador normal de error estándar no es válido.

Normalmente eso lleva a subestimar los errores estándar.

Errores Agrupados II

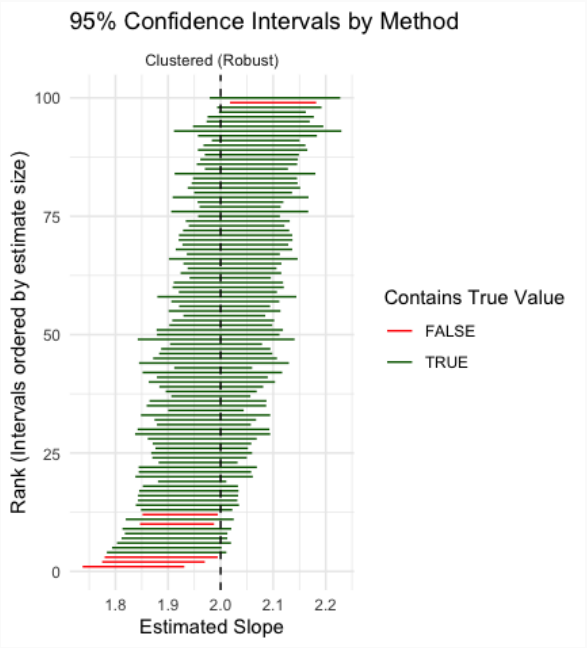
Vamos a entender ese problema con una simulación.


95% Confidence Intervals by Method
No Clustering
Rank (Intervals ordered by estimate size)
100
75
50
25
0
1.90
1.95
2.00
2.05
2.10
Estimated Slope
Contains True Value
FALSE
TRUE


95% Confidence Intervals by Method
Clustered (Naive)
Rank (Intervals ordered by estimate size)
100
75
50
25
0
1.8
1.9
2.0
2.1
2.2
Estimated Slope
Contains True Value
FALSE
TRUE

Errores Agrupados II

Si no llevamos en cuenta la estructura de los errores, la prueba no es válida: rechazamos mucho más que 5% de los casos.


95% Confidence Intervals by Method
Clustered (Robust)
Rank (Intervals ordered by estimate size)
100
75
50
25
0
1.8
1.9
2.0
2.1
2.2
Estimated Slope
Contains True Value
FALSE
TRUE

Errores Agrupados II

Si no conocemos el proceso generador de datos, ¿cómo saber cuál es el nivel apropiado para agrupar?

- Es difícil de obtener una regla general.
- Si tenemos un argumento teórico para agrupar en un cierto nivel, lo hagamos.
- Otra opción es agrupar en el mismo nivel que la variación que se utiliza. Por ej:
  - Si estudiamos escuelas con diferentes practicas de ensino, agrupamos en la escuela.
  - Si las practicas son las mismas para escuelas en una misma comuna, o en un mismo barrio, se agrupa en ese nivel.
- Es particularmente importante cuando tenemos datos de un individuo o un grupo en diferentes momentos en el tiempo.
  - Como una unidad tiende a ser correlacionada con si misma en el tiempo, devemos agrupar en el nivel de la unidad.

Errores Agrupados II

Una limitación de este ajuste es que funciona bien cuando tenemos un número suficiente de grupos (al menos 50).

Si tenemos menos grupos que eso, necesitamos otro tipo de ajuste.

Bootstrap

Bootstrap

Bootstrap es una técnica *extremadamente* útil para hacer inferencia. La ventaja es que el bootstrap funciona *casi siempre*. Y cuando no funciona, normalmente hay una versión un poco diferente que sí funciona.

La idea del bootstrap es que no vamos a intentar modelar la estructura de los datos analíticamente y derivar una fórmula para los errores estándar.

En vez de eso, vamos utilizar los propios datos.

Bootstrap

Idealmente, queremos comprender como nuestra estimación se comporta si la muestra aleatoria que extraímos fuera otra muestra de la misma población.

Pero no tenemos otra muestra de la misma población.

Lo que vamos hacer es utilizar la muestra que tenemos para crear una otra muestra *posible* y ver qué pasa con el estimador. Hacemos eso muchas veces y construímos la distribución del estimador.

El truco es cómo construir esas muestras alternativas.

Bootstrap

Podemos extrair una submuestra de nuestra muestra. Pero eso no va a funcionar bien porque vamos tener menos observaciones.

Si extraímos una muestra de mismo tamaño, sin remplazo, obtenemos una muestra idéntica. Tanpoco sirve.

Pero si extraímos una muestra de mismo tamaño **con remplazo**, vamos tener una muestra un poco diferente.

- Algunas unidades van a estar multiples veces, y otras no van estar.

Bootstrap

El método de bootstrap para estimar el error estándar es:

- 1: Creamos una nueva muestra mediante extración con remplazo de la muestra que tenemos.
- 2: Estimamos nuestro parámetro de interés.
- 3: Repetimos 1 y 2 várias veces (mil o diez mil).
- 4: El error estándar es simplemente la desviación estándar de los estimadores.

Bootstrap

Un ejemplo muy simple:

Imagina que tenemos 4 individuos en la muestra:

i	Y	X
A	4	3
B	2	6
C	10	7
D	6	5

Si hacemos la regresión de Y en X, obtenemos $\beta = 0.97$.

Bootstrap

Una muestra alternativa podría ser:

i	Y	X
C	10	7
C	10	7
D	6	5
B	2	6

En ese caso, estimamos $\beta = 2.55$.

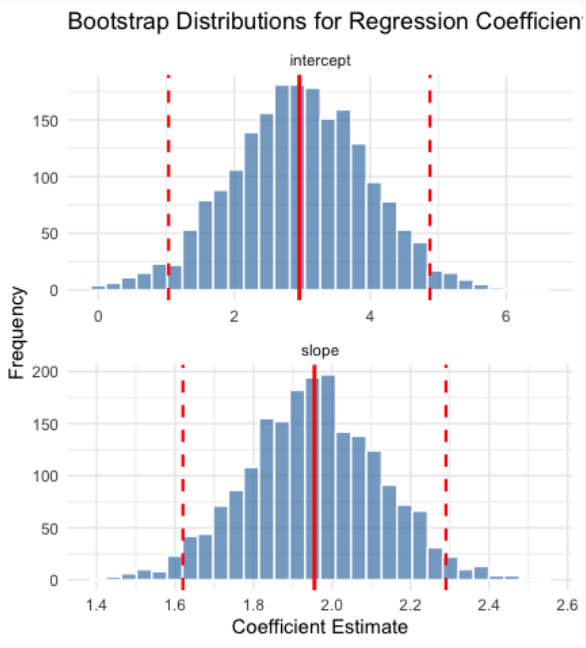
Otra posibilidad:

i	Y	X
A	4	3
A	4	3
B	2	6
B	2	6

Ahora tenemos $\beta = -0.667$.

Si repetimos ese proceso, podemos compreender la variación del estimador.

Vamos ver una demonstración.


Bootstrap Distributions for Regression Coefficien
intercept
150
100
50
0
0
2
4
6
Frequency
slope
200
150
100
50
0
1.4
1.6
1.8
2.0
2.2
2.4
2.6
Coefficient Estimate

Bootstrap

Podemos utilizar el bootstrap de una manera parecida para crear intervalos de confianza:

- 1: Creamos una nueva muestra mediante extración con remplazo de la muestra que tenemos.
- 2: Estimamos nuestro parámetro de interés.
- 3: Repetimos 1 y 2 várias veces (mil o diez mil).
- 4: El intervalo de confianza de 95% es simplemente la parte donde están el 95% de las estimativas.

Bootstrap

Para una regresión simple no hay una gran ventaja de usar el bootstrap, por que conocemos bien cómo estimar los errores estándar.

El bootstrap es más recomendado cuando es difícil de obtener una forma analítica para los errores. Por ejemplo:

- Imaginate que estamos haciendo una estimativa con ajustes como en subclasificación. La estimación incluye calcular várias diferencias y ponderarlas por diferentes pesos. No podemos utilizar los errores normales de una regresión.

Pero podemos escribir el estimador como un código, crear nuevas muestras y utilizar el bootstrap.

Podemos cambiar el bootstrap básico para lidiar con diversos problemas.

Para incorporar errores agrupados con un procedimento de *cluster bootstrap*.

Eso significa que, em vez de extrair nuevas unidades de la muestra com remplazo, una a una, extraímos los grupos enteros.

Por ejemplo: em vez de recrear la muestra por alumno, la recreamos por escuela.

Bootstrap

Un de los pocos casos en que el bootstrap **no funciona bien** es con emparajamiento con vecino más cercano.

La razón es que el bootstrap funciona basado en cambios "suaves" en el estimador: si la muestra cambia un poco, queremos que el estimador cambie un poco.

El caso de emparejamiento con vecino más cercano es uno en que el estimador no cambia para cambios pequeños en la muestra, por que el vecino más cercano tiende a ser todavía el más cercano bajo un poco de variación.

Por esa razón, otros estimadores de emparejamiento son más utilizados, como el kernel (que pondera muchos vecinos cercanos) o el de ponderación de probabilidad inversa.

¿Podemos Confiar en Métodos No Experimentales?

Lalonde (1986)

En su paper **Evaluating the Econometric Evaluations of Training Programs with Experimental Data**, Lalonde se preguntó si es posible utilizar métodos no experimentales para replicar los resultados de ensayos aleatorizados.

Para ello, utilizó datos de un ensayo aleatorizado de capacitación laboral que ocurrió en los años 70.

- Podemos utilizar el grupo de control experimental para obtener resultados sin riesgo de sesgo.
- Después construimos un grupo de comparación con datos observacionales y comparamos los resultados.

El Experimento

El *National Supported Work Demonstration* (NSW) fue un programa que empleó personas vulnerables por entre 9 y 18 meses.

Incluió mujeres en un programa llamado AFDC (*Aid for Families with Dependent Children*), y hombres. Ex-criminales, exadictos, personas que no terminaron el *high school*.

TABLE 1—THE SAMPLE MEANS AND STANDARD DEVIATIONS OF PRE-TRAINING EARNINGS AND OTHER CHARACTERISTICS FOR THE NSW AFDC AND MALE PARTICIPANTS

	Full National Supported Work Sample			
	AFDC Participants		Male Participants	
Variable	Treatments	Controls	Treatments	Controls
Age	33.37	33.63	24.49	23.99
	(7.43)	(7.18)	(6.58)	(6.54)
Years of School	10.30	10.27	10.17	10.17
	(1.92)	(2.00)	(1.75)	(1.76)
Proportion	.70	.69	.79	.80
High School Dropouts	(.46)	(.46)	(.41)	(.40)
Proportion Married	.02	.04	.14	.13
	(.15)	(.20)	(.35)	(.35)
Proportion Black	.84	.82	.76	.75
	(.37)	(.39)	(.43)	(.43)
Proportion Hispanic	.12	.13	.12	.14
	(.32)	(.33)	(.33)	(.35)
Real Earnings	$393	$395	1472	1558
1 year Before	(1,203)	(1,149)	(2656)	(2961)
Training	[43]	[41]	[58]	[63]
Real Earnings	$854	$894	2860	3030
2 years Before	(2,087)	(2,240)	(4729)	(5293)
Training	[74]	[79]	[104]	[113]
Hours Worked	90	92	278	274
1 year Before	(251)	(253)	(466)	(458)
Training	[9]	[9]	[10]	[10]
Hours Worked	186	188	458	469
2 years Before	(434)	(450)	(654)	(689)
Training	[15]	[16]	[14]	[15]
Month of Assignment	−12.26	−12.30	−16.08	−15.91
(Jan. 78 = 0)	(4.30)	(4.23)	(5.97)	(5.89)
Number of Observations	800	802	2083	2193

Note: The numbers shown in parentheses are the standard deviations and those in the square brackets are the standard errors.

Lalonde (1986)

La tabla de equilíbrio parece buena.

Las diferencias son pequeñas y ninguna es estadísticamente significativa.

Lalonde (1986)

Agora vamos mirar los grupos de control no experimentales.

Vienen de dos encuestas: el *Panel Study of Income Dynamics* (PSID) y el *Current Population Survey* (CPS).

Lalonde (1986)

Con ellas, Lalonde crea 6 grupos de comparación:

- PSID 1: hombres jefes de família en PSID entre 1975 y 1978, con menos de 55 años y no jubilados.
- PSID 2: ... que no trabajaban en 76.
- PSID 3: ... que no trabajaban en 75 y 76.
- CPS 1: hombres en CPS con menos de 55 años.
- CPS 2: ... que no trabajaban en 76.
- CPS 3: ... y bajo el nivel de pobreza en 75.

TABLE 3—ANNUAL EARNINGS OF NSW MALE TREATMENTS, CONTROLS, AND SIX CANDIDATE COMPARISON GROUPS FROM THE *PSID* AND *CPS-SSA*

Year	Treatments	Controls	Comparison Group[a,b] *PSID*-1	*PSID*-2	*PSID*-3	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3
1975	$3,066	$3,027	19,056[a]	7,569	2,611	13,650	7,387	2,729
	(283)	(252)	(272)	(568)	(492)	(73)	(206)	(197)
1976	$4,035	$2,121	20,267	6,152	3,191	14,579	6,390	3,863
	(215)	(163)	(296)	(601)	(609)	(75)	(187)	(267)
1977	$6,335	$3,403	20,898	7,985	3,981	15,046	9,305	6,399
	(376)	(228)	(296)	(621)	(594)	(76)	(225)	(398)
1978	$5,976	$5,090	21,542	9,996	5,279	14,846	10,071	7,277
	(402)	(227)	(311)	(703)	(686)	(76)	(241)	(431)
Number of Observations	297	425	2,493	253	128	15,992	1,283	305

[a] The Comparison Groups are defined as follows: *PSID*-1: All male household heads continuously from 1975 through 1978, who were less than 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all men who were not working when surveyed in the spring of 1976; *PSID*-3: Selects from the *PSID*-1 group all men who were not working when surveyed in either spring of 1975 or 1976; *CPS-SSA*-1: All males based on Westat's criteria, except those over 55-years-old; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all males who were not working when surveyed in March 1976; *CPS-SSA*-3: Selects from the *CPS-SSA*-1 unemployed males in 1976 whose income in 1975 was below the poverty level.

Con estos grupos, Lalonde testa vários estimadores:

- La diferencia de medias de ingresos
- La diferencia de medias controlando por edad, edad al cuadrado, educación, si uno terminó el high school, y raza.
- La diferencia de *la evolución de los ingresos*.
- La diferencia de la evolución de los ingresos, con controles.
- La diferencia de medias de ingresos controlando por el nivel de ingresos anterior
- La diferencia de medias de ingresos controlando por el nivel de ingresos anterior y otros controles
- La diferencia de medias de ingresos con todos los controles anteriores posibles.

TABLE 5—EARNINGS COMPARISONS AND ESTIMATED TRAINING EFFECTS FOR THE NSW MALE PARTICIPANTS USING COMPARISON GROUPS FROM THE *PSID* AND THE *CPS-SSA*[a,b]

Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–78 (1)	NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2)	Pre-Training Year, 1975 Adjusted[c] (3)	Post-Training Year, 1978 Unadjusted (4)	Post-Training Year, 1978 Adjusted[c] (5)	Difference in Differences: Difference in Earnings Growth 1975–78 Treatments Less Comparisons: Without Age (6)	With Age (7)	Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–78: Unadjusted (8)	Adjusted[c] (9)	Controlling for All Observed Variables and Pre-Training Earnings (10)
Controls	$2,063	$39	$-21	$886	$798	$847	$856	$897	$802	$662
	(325)	(383)	(378)	(476)	(472)	(560)	(558)	(467)	(467)	(506)
PSID-1	$2,043	-$15,997	-$7,624	-$15,578	-$8,067	$425	-$749	-$2,380	-$2,119	-$1,228
	(237)	(795)	(851)	(913)	(990)	(650)	(692)	(680)	(746)	(896)
PSID-2	$6,071	-$4,503	-$3,669	-$4,020	-$3,482	$484	-$650	-$1,364	-$1,694	-$792
	(637)	(608)	(757)	(781)	(935)	(738)	(850)	(729)	(878)	(1024)
PSID-3	($3,322	($455	$455	$697	-$509	$242	-$1,325	$629	-$552	$397
	(780)	(539)	(704)	(760)	(967)	(884)	(1078)	(757)	(967)	(1103)
CPS-SSA-1	$1,196	-$10,585	-$4,654	-$8,870	-$4,416	$1,714	$195	-$1,543	-$1,102	-$805
	(61)	(539)	(509)	(562)	(557)	(452)	(441)	(426)	(450)	(484)
CPS-SSA-2	$2,684	-$4,321	-$1,824	-$4,095	-$1,675	$226	-$488	-$1,850	-$782	-$319
	(229)	(450)	(535)	(537)	(672)	(539)	(530)	(497)	(621)	(761)
CPS-SSA-3	$4,548	$337	$878	-$1,300	$224	-$1,637	-$1,388	-$1,396	$17	$1,466
	(409)	(343)	(447)	(590)	(766)	(631)	(655)	(582)	(761)	(984)

[a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1978. Based on the experimental data an unbiased estimate of the impact of training presented in col. 4 is $886. The first three columns present the difference between each comparison group's 1975 and 1978 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.

[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.

[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

[d] See Table 3 for definitions of the comparison groups.

Lalonde (1986)

Los resultados para el grupo experimental son *muy similares* dentre todos los estimadores.

Los resultados no experimentales son altamente variables. Mismo los casos con más restricciones y controles no puede replicar los resultados experimentales.

En general, un resultado *preocupante* para el uso de evaluaciones no experimentales.

Dehejia y Wahba (1999)

Ese problema fue reanalizado por Dehejia y Wahba en su paper *Causal effects in nonexerimental studies: reevaluating the evaluation of training programs*.

Su principal contribución es de utilizar métodos de emparejamiento y tener atención al soporte común (*common support*).

- Más cuidado en elegir controles más comparables (no solo em promedio).

Dehejia y Wahba (1999)

Lo primero que hacen es replicar los resultados de Lalonde, y utilizar una muestra con más información sobre los ingresos pasados.

- Excluen los que no tienen información de ingresos en 74.
- También excluen los que entraran en el programa más tarde y tenían ingresos cuando el programa empezó.

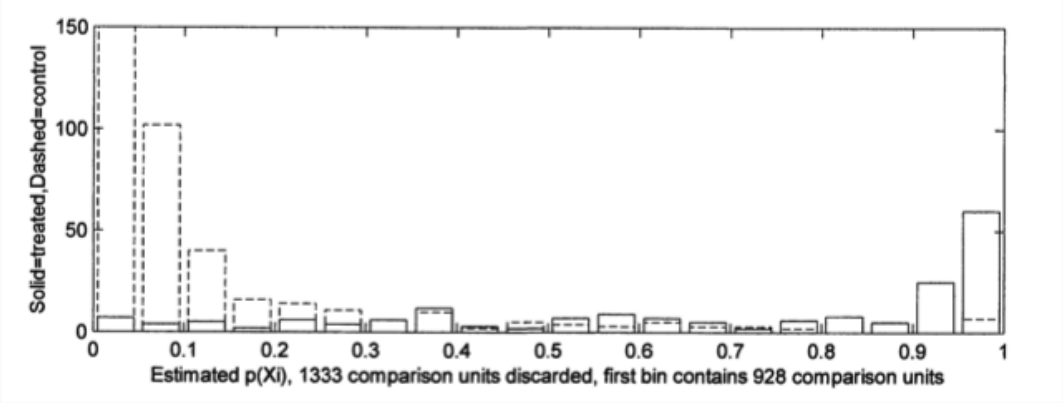
Ellos muestran que los resultados son un poco diferentes, pero la conclusión no cambia solo por la diferencia de muestra.

	A. Lalonde's original sample					C. RE74 subsample (results use RE74)				
	NSW treatment earnings less comparison group earnings 1978		Unrestricted differences in differences: Quasi-difference in earnings growth 1975–1978		Controlling for all variables[f]	NSW treatment earnings less comparison group earnings 1978		Unrestricted differences in differences: Quasi-difference in earnings growth 1975–1978		Controlling for all variables[f]
Comparison group	Unadjusted[b] (1)	Adjusted[c] (2)	Unadjusted[d] (3)	Adjusted[e] (4)	(5)	Unadjusted[b] (1)	Adjusted[c] (2)	Unadjusted[d] (3)	Adjusted[e] (4)	(5)
NSW	886	798	879	802	820	1,794	1,688	1,750	1,672	1,655
	(472)	(472)	(467)	(468)	(468)	(633)	(636)	(632)	(638)	(640)
PSID-1	−15,578	−8,067	−2,380	−2,119	−1,844	−15,205	−879	−582	218	731
	(913)	(990)	(680)	(746)	(762)	(1155)	(931)	(841)	(866)	(886)
PSID-2	−4,020	−3,482	−1,364	−1,694	−1,876	−3,647	94	721	907	683
	(781)	(935)	(729)	(878)	(885)	(960)	(1042)	(886)	(1004)	(1028)
PSID-3	697	−509	629	−552	−576	1,070	821	1,370	822	825
	(760)	(967)	(757)	(967)	(968)	(900)	(1100)	(897)	(1101)	(1104)
CPS-1	−8,870	−4,416	−1,543	−1,102	−987	−8,498	−8	−78	739	972
	(562)	(577)	(426)	(450)	(452)	(712)	(572)	(537)	(547)	(550)
CPS-2	−4,195	−2,341	−1,649	−1,129	−1,149	−3,822	615	−263	879	790
	(533)	(620)	(459)	(551)	(551)	(671)	(672)	(574)	(654)	(658)
CPS-3	−1,008	−1	−1,204	−263	−234	−635	1,270	−91	1,326	1,326
	(539)	(681)	(532)	(677)	(675)	(657)	(798)	(641)	(796)	(798)

En seguida, estiman el propensity score para cada grupo de comparación y eliminan controles fuera del soporte común.

Los modelos para estimar el PS son modelos logit que incluyen edad, edad al cuadrado, educación, educación al cuadrado, casado, no tiene diploma, raza, ingresos en 74 y 75 y sus cuadrados, desempleo en 74 y 75. Además, también incluyen:

- PSID: desempleo en 74 * negro.
- CPS: educación * ingresos en 74, y edad al cubo.


Solid=treated,Dashed=control
150
100
50
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Estimated p(Xi), 1333 comparison units discarded, first bin contains 928 comparison units

Con esta función de PS, ellos utilizan 7 estimadores:

- La diferencia simple
- La diferencia con controles
- La diferencia controlando por el propensity score (y su cuadrado)
- Con estratificación por valores del PS (subclasificación)
- Con estratificación por valores del PS (subclasificación) más controles
- Con matching en el vecino más próximo por el PS
- Con matching en el vecino más próximo por el PS más controles

Table 3. Estimated Training Effects for the NSW Male Participants Using Comparison Groups From PSID and CPS

	NSW earnings less comparison group earnings		NSW treatment earnings less comparison group earnings, conditional on the estimated propensity score					
			Quadratic in score[b]	Stratifying on the score			Matching on the score	
	(1) Unadjusted	(2) Adjusted[a]	(3)	(4) Unadjusted	(5) Adjusted	(6) Observations[c]	(7) Unadjusted	(8) Adjusted[d]
NSW	1,794	1,672						
	(633)	(638)						
PSID-1[e]	−15,205	731	294	1,608	1,494	1,255	1,691	1,473
	(1,154)	(886)	(1,389)	(1,571)	(1,581)		(2,209)	(809)
PSID-2[f]	−3,647	683	496	2,220	2,235	389	1,455	1,480
	(959)	(1,028)	(1,193)	(1,768)	(1,793)		(2,303)	(808)
PSID-3[f]	1,069	825	647	2,321	1,870	247	2,120	1,549
	(899)	(1,104)	(1,383)	(1,994)	(2,002)		(2,335)	(826)
CPS-1[g]	−8,498	972	1,117	1,713	1,774	4,117	1,582	1,616
	(712)	(550)	(747)	(1,115)	(1,152)		(1,069)	(751)
CPS-2[g]	−3,822	790	505	1,543	1,622	1,493	1,788	1,563
	(670)	(658)	(847)	(1,461)	(1,346)		(1,205)	(753)
CPS-3[g]	−635	1,326	556	1,252	2,219	514	587	662
	(657)	(798)	(951)	(1,617)	(2,082)		(1,496)	(776)

Los estimadores con propensity score tienen resultados mucho mejores.

- Un detalle es que sus errores son muy grandes, entonces así mismo no estaríamos seguros del efecto.

En general, un resultado muy positivo para los métodos no experimentales.

- Motivó bastante interés en esos métodos.

Pero, hubo una re-re-evaluación.

Smith and Todd (2005)

Smith and Todd revisitan ese debate en su paper: **Does matching overcome LaLonde's critique of non-experimental estimators?**

Esencialmente, su contribución es testar si los métodos de propensity score son *siempre* mejores, o si ciertas elecciones de WB influenciaran sus resultados.

Smith and Todd (2005)

Su primero resultado importante es aplicar los métodos de PS de WB para (1) la muestra original de Lalonde, y (2) una muestra sin una restrición que WB utilizaran (eliminando individuos que empezaran más tarde con ingresos positivos antes del programa).

Eses resultados no tienen el pequeño sesgo que WB encontraron. Pelo contrario, tienen un sesgo muy grande.

Conclusión: métodos de PS pueden ser afetados por pequeños cambios en la muestra, y no siempre funcionan tan bien.

Smith and Todd (2005)

Segundo, Smith and Todd utilizan la misma muestra que WB, pero construyen el PS con solamente las variables que tenía Lalonde.

Otra vez, los resultados son malos. El sesgo es grande, y los resultados varían muchísimo con cambios en el estimador.

Conclusión: métodos de PS también pueden ser afectados por exactamente cuales variables están disponibles.

Smith and Todd (2005)

Tercero, Smith and Todd implementan otros estimadores de PS con aún más estructura. En particular, un estimador que mezcla PS con DID (que vamos estudiar más adelante).

Ese estimador resulta un poco mejor, pero todavía tiene problemas.

Smith and Todd (2005)

Finalmente, Smith and Todd muestran que, cuando se utiliza la misma muestra que WB, y se utiliza métodos más simples, como regresiones con los mismos controles, los resultados son bastante buenos.

Conclusión: la mejora en los resultados de WB fue más por causa de sus restricciones en la muestra que por utilizar métodos de matching.

Basicamente, cuando WB excluyeron personas sin ingresos pasados, seleccionaran una muestra con un problema de sesgo *más fácil* de solucionar.

Smith and Todd (2005)

La conclusión general es que propensity score no es una solución general.

Cuando tenemos mucha información y de buena calidad, métodos simples resultan bien. Cuándo tenemos pocas variables, y los resultados son medidos de manera inconsistente, incluso métodos avanzados encuentran problemas.

Importante compreender los datos y sus problemas, y elegir métodos apropiados.