El supuesto de Independencia Condicional

Ricardo Pasquini

Mayo 2024

Sesgo en diferencias post-tratamiento

- Anteriormente, obtuvimos una expresión para el resultado de tomar la diferencia ex-post entre afectados por una intervención o programa de los no afectados.
- Esta diferencia era igual al Tratamiento sobre los Tratados más un sesgo de selección:

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \text{ATT} + \text{Preexistent bias}$$

Sesgo en diferencias post-tratamiento

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1] + E[Y_{0i}|T_i =$$

CIA - Selection on Observables

- El Supuesto de Independencia Condicional (CIA) afirma que, condicionalmente a una serie de características observadas X_i, los resultados potenciales son independientes del tratamiento.
- Esto es:
$$Y_{0i}, Y_{1i} \perp D_i \mid X_i$$
- O de forma equivalente:

$$E[Y_{0i}|X_i, T_i = 1] = E[Y_{0i}|X_i, T_i = 0]$$

CIA - Selection on Observables

- Por lo tanto, el sesgo de selección desaparece y la diferencia entre participantes y no participantes dará como resultado el efecto del tratamiento:

$$E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0] = E[Y_{1i} - Y_{0i}|T_i = 1]$$

Matching

- ¿Cómo podemos recuperar el ATT si se cumple CIA?

  $\delta_{ATT} = E[Y_{1i} - Y_{0i} | T_i = 1] = E[E[Y_i | X_i, T_i = 1] - E[Y_i | X_i, T_i = 0] |$

- X CIA:

$$E[Y_{0i} | T_i = 0] = E[Y_{0i} | T_i = 1]$$

Matching

$$\delta_{ATT} = E[E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0]|T_i = 1] = E[\delta_X|T_i = 1]$$

donde

$$\delta_X = E_Y[y_{1i}|X_i, T_i = 1] - E_Y[y_{0i}|X_i, T_i = 0]$$

Matching. El estimador

- En el caso que X toma valores discretos vale que:

$$ATT = E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1] = \sum_x \delta_x P(X_i = x|T_i = 1)$$

- Por lo tanto, un estimador para el ATT es:

$$A\hat{T}T = \sum_x \hat{\delta}_x \hat{P}(X_i = x|T_i = 1)$$

CASO: Dale y Krueger 2002

ESTIMATING THE PAYOFF TO ATTENDING A MORE SELECTIVE COLLEGE: AN APPLICATION OF SELECTION ON OBSERVABLES AND UNOBSERVABLES*

STACY BERG DALE AND ALAN B. KRUEGER

Estimates of the effect of college selectivity on earnings may be biased because elite colleges admit students, in part, based on characteristics that are related to future earnings. We matched students who applied to, and were accepted by, similar colleges to try to eliminate this bias. Using the College and Beyond data set and National Longitudinal Survey of the High School Class of 1972, we find that students who attended more selective colleges earned about the same as students of seemingly comparable ability who attended less selective schools. Children from low-income families, however, earned more if they attended selective colleges.

Figure: Dale y Krueger

CASO: Dale y Krueger 2002

Past studies have found that students who attended colleges with higher average SAT scores or higher tuition tend to have higher earnings when they are observed in the labor market. Attending a college with a 100 point higher average SAT is associated with 3 to 7 percent higher earnings later in life (see, e.g., Kane [1998]). As Kane notes, an obvious concern with this conclusion is that students who attend more elite colleges may have greater earnings capacity regardless of where they attend school.

CASO: Dale y Krueger 2002 (simplificado)

TABLE 2.1
The college matching matrix

		Private			Public			
Applicant group	Student	Ivy	Leafy	Smart	All State	Tall State	Altered State	1996 earnings
A	1		Reject	Admit		Admit		110,000
	2		Reject	Admit		Admit		100,000
	3		Reject	Admit		Admit		110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Note: Enrollment decisions are highlighted in gray.

Figure: College Matrix

CASO: Dale y Krueger 2002 (simplificado)

- La diferencia de privada vs pública me da aprox 20,000 (92 vs 72.5 en promedio)
- La diferencia si matcheamos por grupo y ponderamos es:

$$\frac{3}{5}(-5,000) + \frac{2}{5}(30,000) = 9,000$$

- Notar que al matchear tengo que descartar los grupos C y D

CASO: Dale y Krueger 2002 (simplificado)

TABLE I
ILLUSTRATION OF HOW MATCHED-APPLICANT GROUPS WERE CONSTRUCTED

		Student applications to college							
		Application 1		Application 2		Application 3		Application 4	
Student	Matched-applicant group	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision
Student A	1	1280	Reject	1226	Accept*	1215	Accept	na	na
Student B	1	1280	Reject	1226	Accept	1215	Accept*	na	na
Student C	2	1360	Accept	1310	Reject	1270	Accept*	1155	Accept
Student D	2	1355	Accept	1316	Reject	1270	Accept*	1160	Accept
Student E	2	1370	Accept*	1316	Reject	1260	Accept	1150	Accept
Student F	Excluded	1180	Accept*	na	na	na	na	na	na
Student G	Excluded	1180	Accept*	na	na	na	na	na	na
Student H	3	1360	Accept	1308	Accept*	1260	Accept	1160	Accept
Student I	3	1370	Accept*	1311	Accept	1255	Accept	1155	Accept
Student J	3	1350	Accept	1316	Accept*	1265	Accept	1155	Accept
Student K	4	1245	Reject	1217	Reject	1180	Accept*	na	na
Student L	4	1235	Reject	1209	Reject	1180	Accept*	na	na
Student M	5	1140	Accept	1055	Accept*	na	na	na	na
Student N	5	1145	Accept*	1060	Accept	na	na	na	na
Student O	No match	1370	Reject	1038	Accept*	na	na	na	na

* Denotes school attended.
na = did not report submitting application.
The data shown on this table represent hypothetical students. Students F and G would be excluded from the matched-applicant subsample because they applied to only one school (the school they attended). Student O would be excluded because no other student applied to an equivalent set of institutions.

Digresión: CIA y CEF

- Notar que una manera alternativa de estimar:

$$E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0]$$

  es apelando al modelo de regresión. Cuando vimos podemos implementar un CEF perfectamente cuando las categorías de la variable X son finitas.

- En este caso, el supuesto sería:

$$E[Y_i|X_i, T_i] = \beta_0 + \beta_1 X + \beta_2 T + \epsilon$$

- Si no tenemos categorías finitas, igualmente debiéramos encontrar resultados similares. La diferencia entre los dos métodos es que la regresión pondera las observaciones de acuerdo a la varianza relativa, mientras el matching de acuerdo al número de observaciones (Véase MHE.)

CASO: Dale y Krueger 2002 (simplificado)

- En el caso de Dale y Krueger, el modelo de regresión implementado es:

$$\log(\text{earnings}) = \alpha_i + \beta_1 \text{Private}_i + \beta_2 \text{SAT} + \epsilon_i$$

CASO: Dale y Krueger 2002 (simplificado)

TABLE 2.2
Private school effects: Barron's matches

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135 (.055)	.095 (.052)	.086 (.034)	.007 (.038)	.003 (.039)	.013 (.025)
Own SAT score ÷ 100		.048 (.009)	.016 (.007)		.033 (.007)	.001 (.007)
Log parental income			.219 (.022)			.190 (.023)
Female			−.403 (.018)			−.395 (.021)
Black			.005 (.041)			−.040 (.042)
Hispanic			.062 (.072)			.032 (.070)
Asian			.170 (.074)			.145 (.068)
Other/missing race			−.074 (.157)			−.079 (.156)
High school top 10%			.095 (.027)			.082 (.028)
High school rank missing			.019 (.033)			.015 (.037)
Athlete			.123 (.025)			.115 (.027)
Selectivity-group dummies	No	No	No	Yes	Yes	Yes

Control for Covariates Using the Propensity Score

- Propensity Score Matching (PSM) construye un grupo de comparación estadística que se basa en un modelo de probabilidad de participar en el tratamiento, utilizando características observadas.
- La idea es emparejar a cada individuo del grupo de tratamiento con un individuo del grupo de control con una probabilidad similar de participar en el programa.

Propensity Score Matching (PSM)

Esto implica dos supuestos:

- CIA: Todo lo que puede explicar la condición de tratamiento se debe a factores observables.
- Presencia de un soporte común o superposición, que garantice que podré encontrar un individuo “similar” en el grupo de control. Esto es que:

$$0 < P(T_i = 1|X_i) < 1$$

Propensity Score Matching (PSM)

- Una forma de calcular esto es utilizando una regresión logística (Logit) o una regresión Probit.

$$P(T_i = 1|X_i) = F(X_i\beta)$$

- Luego podemos usar esta probabilidad para hacer matching:

$$ATT = E[Y_{1i} - Y_{0i}|T_i = 1] = E[E[Y_{1i} - Y_{0i}|X_i]|T_i = 1]$$