

Matching

PP5001 · Week 2

Professor David A. Jaeger

From Random Assignment to Selection on Observables

Last week, random assignment gave us a credible comparison group. What can we do when participation is **not randomly assigned**?

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \underbrace{E[Y_i^1 - Y_i^0 | D_i = 1]}_{\text{ATT}} + \underbrace{E[Y_i^0 | D_i = 1] - E[Y_i^0 | D_i = 0]}_{\text{selection bias}}.$$

Matching aims to make an “**apples to apples**” **comparison** using observed characteristics. Whether it removes selection bias depends on what we observe and how participation is determined.

The Supported Work Experiment

For discussion

Impacts for Women: Table 4-3

Table 4-3. Average Earnings per Month, by Length of Follow-Up: AFDC Sample

<i>Month After Enrollment in the Sample</i>	<i>Experimental Group Mean</i>	<i>Control Group Mean</i>	<i>Experimental-Control Differential</i>	<i>Mean Earnings From Supported Work^a</i>	
				<i>Earnings</i>	<i>As Percent of Total Earnings of Experimentals</i>
1-3	\$409.44	\$49.28	\$360.16†	\$401.27	98.0
4-6	412.30	90.84	321.46†	389.01	94.4
7-9	387.20	99.28	287.92†	348.51	90.0
10-12	371.25	117.85	253.40†	275.45	74.2
13-15	237.11	140.12	96.99†	65.76	27.7
16-18	213.30	159.33	53.97†	13.99	6.6
19-21	230.92	161.79	69.13†	10.97	4.8
22-24	244.64	166.54	78.10†	2.93	1.2
25-27	248.47	167.86	80.61†	0.00	0.0

Source: See Table 4-1.

Notes: The data presented in the first three columns are regression-adjusted estimates that control for differences of age, sex, race, education, prior work experience, household composition, site, and length of site operation. Averages are calculated for all members of the sample, including those with no employment in the covered period. Dollar amounts reported are gross earnings unadjusted for inflation.

^aThese data are not regression adjusted. No experimentals should have been in the program beyond month 21. The small number indicated as participating after that time reflects either data errors or failures by program operators to terminate individuals on schedule.

†Statistically significant at the 5 percent level.

Ex-Addict Sample: Table 5-3

Table 5-3. Average Earnings Per Month, by Length of Follow-Up: Ex-Addict Sample

<i>Month After Enrollment in the Sample</i>	<i>Experimental Group Mean</i>	<i>Control Group Mean</i>	<i>Experimental-Control Differential</i>	<i>Mean Earnings from Supported Work^a</i>	
				<i>Earnings</i>	<i>As Percent of Total Earnings of Experimentals</i>
1-3	\$395.31	\$122.30	\$273.01†	\$355.31	89.9
4-6	348.00	184.58	163.42†	284.68	81.8
7-9	306.08	166.99	139.09†	224.83	73.5
10-12	280.71	205.37	75.34†	154.02	54.9
13-15	251.65	211.81	39.84*	66.17	26.3
16-18	215.51	222.22	-6.71	17.27	8.0
19-21	243.78	250.25	-6.47	5.20	2.1
22-24	281.02	270.36	10.66	3.81	1.4
25-27	287.05	259.88	27.17	1.78	0.6
28-30	304.09	237.45	66.64	0.00	0.0
31-33	332.18	221.85	110.33†	0.00	0.0
34-36	318.60	218.76	99.84†	0.00	0.0

Source: See Table 5-1.

Notes: The data presented in the first three columns are regression-adjusted estimates that control for differences of age, sex, race, education, prior work experience, household composition, site, and length of site operation. Averages are calculated for all members of the sample, including those with no employment in the covered period. Dollar amounts reported are gross earnings unadjusted for inflation.

^aThese data are not regression adjusted. No experimentals should have been in the program beyond month 21. The small number indicated as participating after that time reflects either data errors or failures by program operators to terminate individuals on schedule.

†Statistically significant at the 5 percent level.

*Statistically significant at the 10 percent level.

Youth Sample: Table 6-3

Table 6-3. Average Earnings per Month, by Length of Follow-Up: Youth Sample

<i>Month After Enrollment in the Sample</i>	<i>Experimental Group Mean</i>	<i>Control Group Mean</i>	<i>Experimental- Control Differential</i>	<i>Mean Earnings From Supported Work^a</i>	
				<i>Earnings</i>	<i>As Percent of Total Earnings of Experimentals</i>
1-3	\$389.52	\$100.15	\$289.37†	\$353.90	90.9
4-6	340.76	140.51	200.25†	266.80	78.3
7-9	284.40	138.24	146.16†	197.20	69.3
10-12	255.24	163.51	91.73†	134.10	52.5
13-15	218.79	211.16	7.63	63.33	28.9
16-18	208.78	213.04	-4.26	28.23	13.5
19-21	246.22	220.85	25.37	9.33	3.8
22-24	270.77	256.01	14.76	7.75	2.9
25-27	265.98	268.05	-2.07	1.63	0.6
28-30	300.95	323.53	-22.58	0.00	0.0
31-33	323.60	347.73	-24.13	0.00	0.0
34-36	287.13	331.59	-44.46	0.00	0.0

Source: See Table 6-1.

Notes: The data presented in the first three columns are regression-adjusted estimates that control for differences of age, sex, race, education, prior work experience, household composition, site, and length of site operation. Averages are calculated for all members of the sample, including those with no employment in the covered period. Dollar amounts reported are gross earnings unadjusted for inflation.

^aThese data are not regression adjusted. No experimentals should have been in the program beyond month 21. The small number indicated as participating after that time reflects either data errors or failures by program operators to terminate individuals on schedule.

†Statistically significant at the 5 percent level.

Ex-Offender Sample: Table 7-3

Table 7-3. Average Earnings per Month, by Length of Follow-Up: Ex-Offender Sample

<i>Month After Enrollment in the Sample</i>	<i>Experimental Group Mean</i>	<i>Control Group Mean</i>	<i>Experimental- Control Differential</i>	<i>Mean Earnings From Supported Work^a</i>	
				<i>Earnings</i>	<i>As Percent of Total Earnings of Experimentals</i>
1-3	\$446.91	\$138.74	\$308.17†	\$383.82	85.9
4-6	385.74	185.15	200.59†	265.53	68.8
7-9	323.94	173.74	150.20†	191.17	59.0
10-12	276.56	221.37	55.19†	107.24	38.8
13-15	274.88	246.15	28.73	40.98	14.9
16-18	269.11	250.10	19.01	13.62	5.1
19-21	265.73	252.92	12.81	11.17	4.2
22-24	273.81	265.63	8.18	6.27	2.3
25-27	266.80	263.72	3.08	2.69	1.0
28-30	353.14	285.48	67.66	0.00	0.0
31-33	375.77	319.99	55.78	0.00	0.0
34-36	361.28	295.35	65.93	0.00	0.0

Source: See Table 7-1.

Notes: The data presented in the first three columns are regression-adjusted estimates that control for differences of age, sex, race, education, prior work experience, household composition, site, and length of site operation. Averages are calculated for all members of the sample, including those with no employment in the covered period. Dollar amounts reported are gross earnings unadjusted for inflation.

^aThese data are not regression adjusted. No experimentals should have been in the program beyond month 21. The small number indicated as participating after that time reflects either data errors or failures by program operators to terminate individuals on schedule.

†Statistically significant at the 5 percent level.

Comparing the Results

For discussion

What Is LaLonde Doing?

For discussion

Data

For discussion

LaLonde: Table 3

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

^aThe 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.

^bAll earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. The number of observations refer only to 1975 and 1978. In the other years there are fewer observations. The sample of treatments is smaller than the sample of controls because treatments still in Supported Work as of January 1978 are excluded from the sample, and in the young high school target group there were by design more controls than treatments.

LaLonde: Table 5

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

^aThe 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.

^bEstimates are in 1982 dollars. The numbers in parentheses are the standard errors.

^cThe exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

^dSee Table 3 for definitions of the comparison groups.

Why Do the Experimental Estimates Differ?

For discussion

LaLonde on Follow-up Timing

“First, the partial sample contains a much larger proportion of males who completed a 36 month interview than does the full sample. Second, the training effects for the full sample are much larger for the 36 month interview than for the 27 month interview.”

LaLonde (1984), Chapter 2, footnote 5, pp. 81–82.

LaLonde on Sample Selection

“Thus the process of selecting individuals with complete earnings in 1975 and 1978 caused us to select treatments, but not controls, with higher earnings after the 27th and 36th month interviews. Despite this potential selection bias, this paper treats the partial sample as if it was an experimental data set.”

LaLonde (1984), Chapter 2, footnote 5, p. 82.

What Makes a Comparison “Apples to Apples”?

People who enter a training programme may differ from other workers in their schooling, work history, and previous earnings.

Let X_i collect **pre-treatment characteristics**. We want nonparticipants with characteristics comparable to those of the participants.

This is more demanding than finding people with the same average age. We need comparable combinations of characteristics—and an argument that the remaining differences do not determine untreated outcomes.

Matching cannot repair an important confounder that we have not measured.

The Assumptions for the ATT

Selection on observables (unconfoundedness):

$$Y_i^0 \perp D_i \mid X_i.$$

Given X_i , treatment status provides no further information about the outcome that would occur **without treatment**.

Common support: for every combination of X_i among treated individuals, there must be potential controls:

$$P(D_i = 0 \mid X_i) > 0.$$

We also retain the consistency and no-interference assumptions discussed in Week 1. Their credibility depends on the study design and setting.

Exact Matching

The simplest form of matching compares individuals with the **exact same set of observable characteristics**.

Under unconfoundedness, compare average outcomes **within the cells defined by the X 's**:

$$\hat{\tau}(x) = \overline{Y}_{D=1, X=x} - \overline{Y}_{D=0, X=x}.$$

For example, compare participants and nonparticipants with the same schooling and pre-programme employment history.

The control mean in each cell supplies the missing untreated outcome for treated individuals in that cell.

Averaging the Comparisons

For the **ATT**, average the within-cell differences using the distribution of characteristics **among treated individuals**.

Cell	Share of treated	Treated mean	Control mean	Difference
$X = a$	0.75	6,000	5,000	1,000
$X = b$	0.25	8,000	6,000	2,000

$$\hat{\tau}_{\text{ATT}} = 0.75 (1,000) + 0.25 (2,000) = 1,250.$$

This illustrative calculation weights each treated person's comparison equally. Using a different population's cell shares would target a different average effect.

Why Exact Matching Becomes Difficult

Exact matching is extremely data intensive.

- With several characteristics, the number of combinations grows quickly: ten binary characteristics produce $2^{10} = 1,024$ cells.
- With continuous covariates, such as previous earnings, exact matches may be rare.
- Some cells contain treated individuals but no controls.

Adding relevant characteristics can make the assumption more plausible while making comparable observations harder to find.

The Propensity Score

The **propensity score** is the probability of receiving treatment conditional on the observed characteristics:

$$p(X_i) = P(D_i = 1 \mid X_i).$$

X_i is a vector; $p(X_i)$ is a **single number**.

We will use **logit** to estimate this probability from pre-treatment characteristics.

The purpose is to make the groups comparable on observed characteristics.

Estimating a Probability: LPM and Logit

Let D_i indicate university attendance and X_i measure family income.

$$\text{LPM: } p_i = \beta_0 + \beta_1 X_i,$$

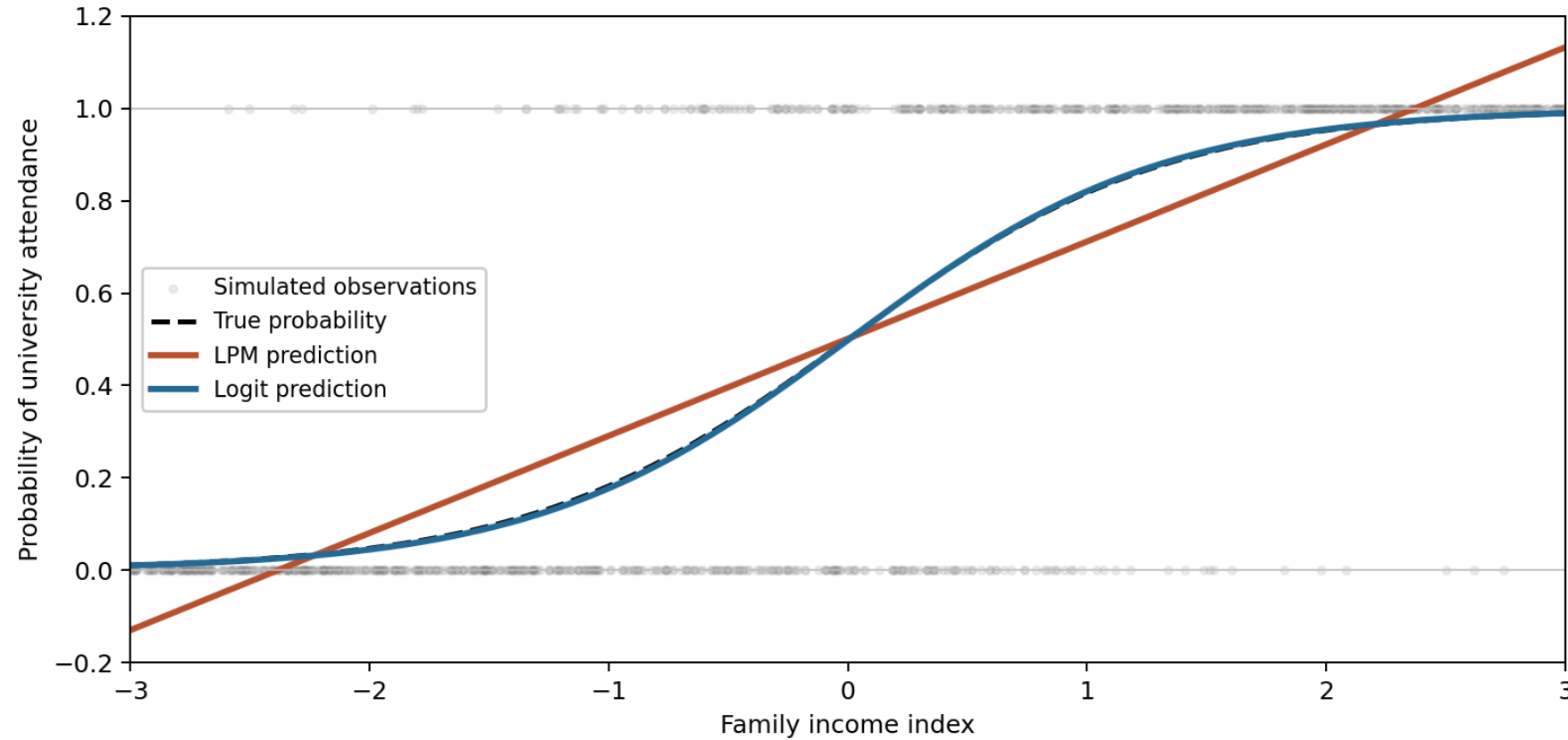
$$\text{Logit: } p_i = \frac{1}{1 + \exp(-(\gamma_0 + \gamma_1 X_i))}.$$

Logit transforms the linear expression into a probability between zero and one.

Latent-variable interpretation: $D_i^* = \gamma_0 + \gamma_1 X_i + \epsilon_i$; attendance occurs when $D_i^* > 0$.

If ϵ_i is independent of X_i and standard logistic, this produces the logit probability above.

LPM and Logit: A Simulated Comparison



Fit with `smf.ols()` or `smf.logit()`, then use `.predict()` to obtain fitted probabilities. The notes give the full simulation and estimation code.

Rosenbaum and Rubin: Dimension Reduction

Rosenbaum and Rubin (1983) show that

$$(Y_i^0, Y_i^1) \perp D_i \mid X_i \implies (Y_i^0, Y_i^1) \perp D_i \mid p(X_i).$$

The key balancing property is

$$D_i \perp X_i \mid p(X_i).$$

If conditioning on X_i makes assignment unconfounded, conditioning on the **true propensity score** does too. This reduces the dimension of the adjustment under the original assumption.

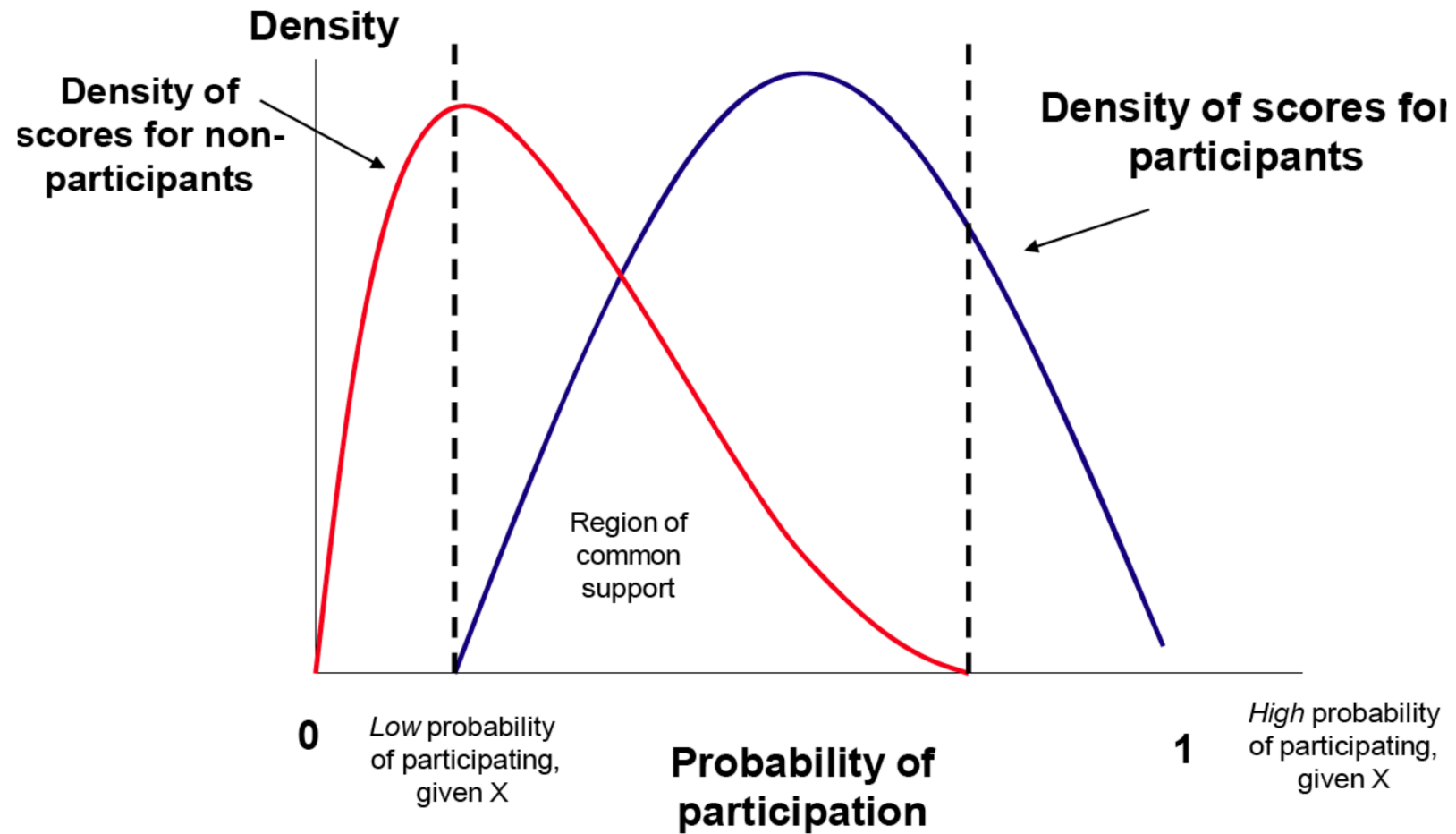
The True Score and the Estimated Score

For the **true** score, balancing is a theorem. For an **estimated** score, balance is something we must check.

The balancing result concerns the **distribution of characteristics** in the treated and control groups conditional on the score.

After matching or weighting, examine whether the observed characteristics are actually balanced.

Common Support



When Comparable Controls Are Scarce

For the ATT, we need controls wherever the treated observations are.

When $p(X_i)$ is close to one, there may be many treated individuals and very few comparable controls. The estimate can depend heavily on those few controls.

Excluding treated observations without support changes the population whose effect we estimate. The estimate then applies to the retained group.

Credible comparisons require observations with comparable characteristics.

Dehejia and Wahba: The Sample

Dehejia and Wahba (1999) reanalyse the **male sample**.

They use the subsample with 1974 earnings available: **185 treated men and 260 experimental controls**, compared with LaLonde's 297 and 425. This gives two years of pre-programme earnings, 1974 and 1975.

Their unadjusted experimental benchmark is **\$1,794**, rather than LaLonde's **\$886**.

PS1 uses this subsample throughout: reproduce its experimental benchmark, then compare IPW using one versus two years of pre-treatment earnings.

Figure 1

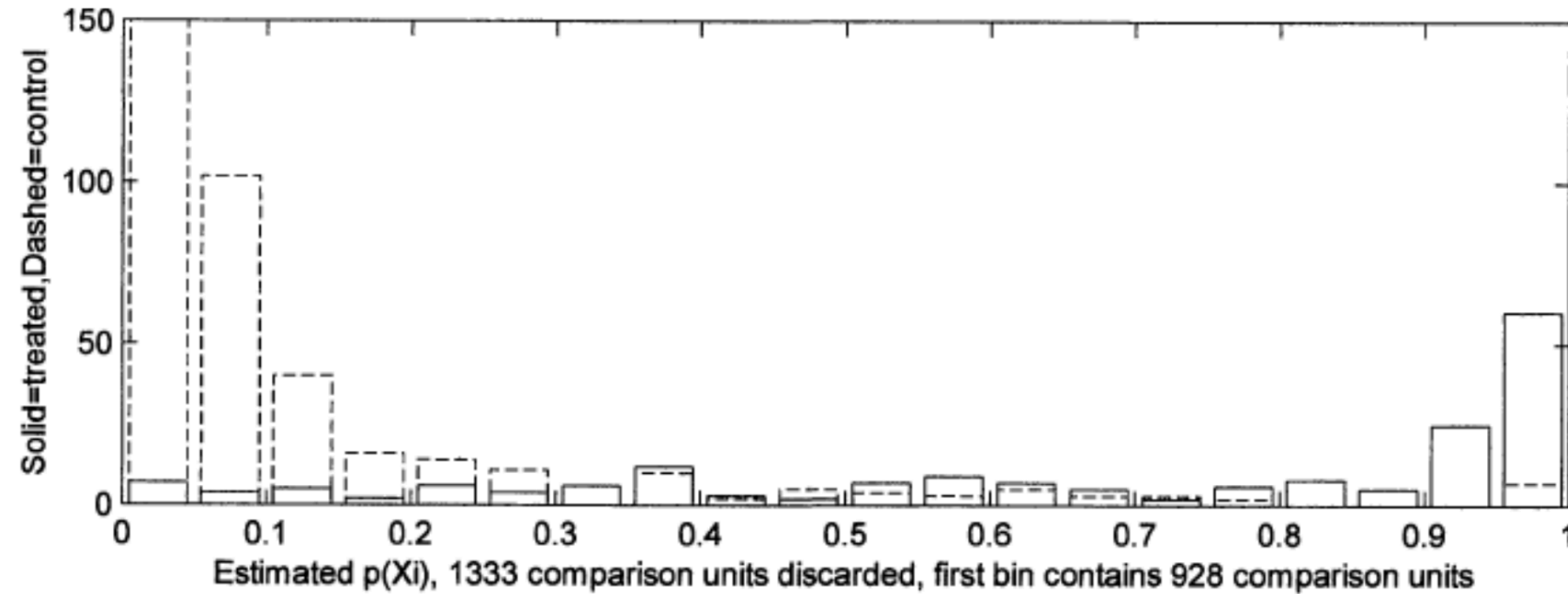


Figure 1. Histogram of the Estimated Propensity Score for NSW Treated Units and PSID Comparison Units. The 1,333 PSID units whose estimated propensity score is less than the minimum estimated propensity score for the treatment group are discarded. The first bin contains 928 PSID units. There is minimal overlap between the two groups. Three bins (.8–.85, .85–.9, and .9–.95) contain no comparison units. There are 97 treated units with an estimated propensity score greater than .8 and only 7 comparison units.

Figure 2

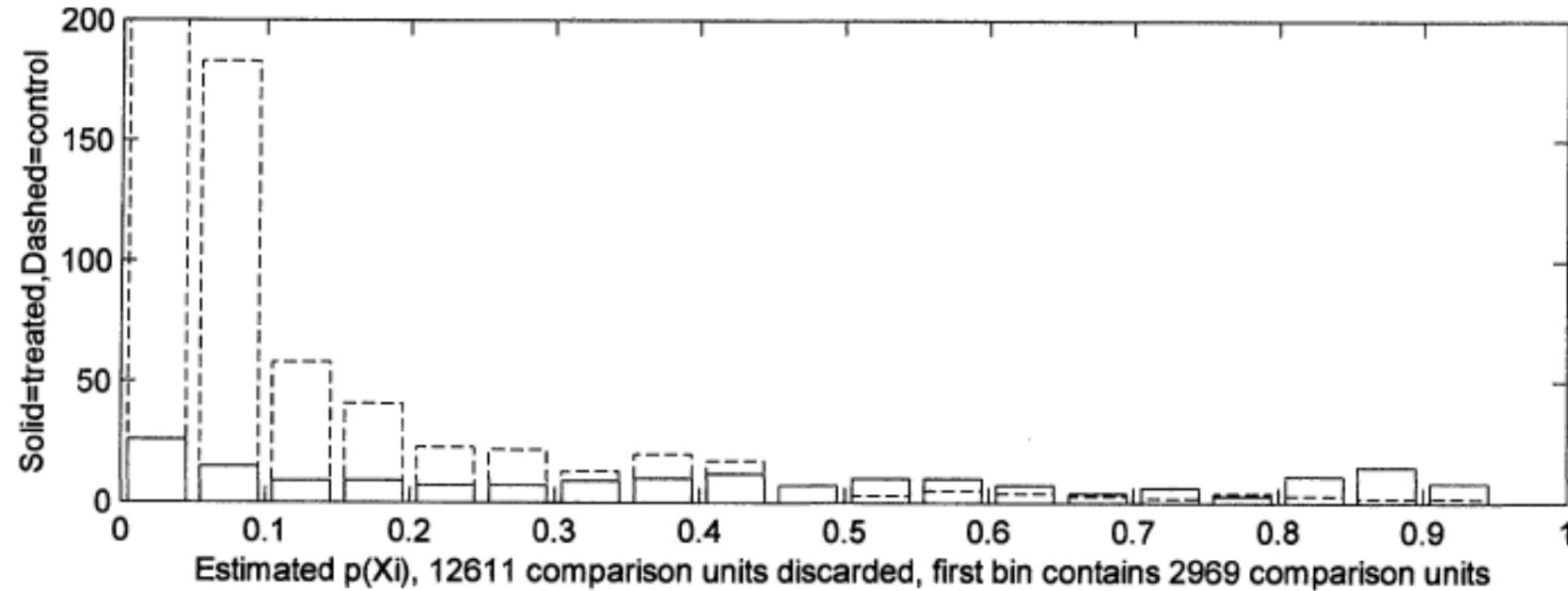


Figure 2. Histogram of the Estimated Propensity Score for NSW Treated Units and CPS Comparison Units. The 12,611 CPS units whose estimated propensity score is less than the minimum estimated propensity score for the treatment group are discarded. The first bin contains 2,969 CPS units. There is minimal overlap between the two groups, but the overlap is greater than in Figure 1; only one bin (.45–.5) contains no comparison units, and there are 35 treated and 7 comparison units with an estimated propensity score greater than .8.

Dehejia and Wahba: Table 3

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

^a Least squares regression: RE78 on a constant, a treatment indicator, age, age², education, no degree, black, Hispanic, RE74, RE75.

^b Least squares regression of RE78 on a quadratic on the estimated propensity score and a treatment indicator, for observations used under stratification; see note (g).

^c Number of observations refers to the actual number of comparison and treatment units used for (3)–(5); namely, all treatment units and those comparison units whose estimated propensity score is greater than the minimum, and less than the maximum, estimated propensity score for the treatment group.

^d Weighted least squares: treatment observations weighted as 1, and control observations weighted by the number of times they are matched to a treatment observation [same covariates as (a)].

Propensity scores are estimated using the logistic model, with specifications as follows:

^e PSID-1: Prob ($T_i = 1$) = F(age, age², education, education², married, no degree, black, Hispanic, RE74, RE75, RE74², RE75², u74*black).

^f PSID-2 and PSID-3: Prob ($T_i = 1$) = F(age, age², education, education², no degree, married, black, Hispanic, RE74, RE74², RE75, RE75², u74, u75).

^g CPS-1, CPS-2, and CPS-3: Prob ($T_i = 1$) = F(age, age², education, education², no degree, married, black, Hispanic, RE74, RE75, u74, u75, education*RE74, age³).

What Should We Learn from the Comparison?

Propensity-score adjustment can substantially change the estimates. Assess unconfoundedness using the assignment process and available characteristics in each application.

The **data comparability** lesson from your reading matters:

- Do the groups face comparable labour markets?
- Are outcomes and covariates measured in the same way?
- Do we observe the characteristics that jointly predict participation and outcomes?

Good balance on measured characteristics is useful evidence about the adjustment. It cannot establish balance on unmeasured confounders.

From Matching to Weighting

Instead of selecting particular controls as matches, we can give controls different weights.

For the **ATT**, retain the treated sample and reweight the controls to resemble it on observed characteristics.

Controls whose characteristics make treatment more likely receive greater weight. Controls who look unlike the treated receive less weight.

This is **propensity-score weighting**, using ATT weights.

ATT Weights

$$w_i = \begin{cases} 1 & \text{if } D_i = 1, \\ \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} & \text{if } D_i = 0. \end{cases}$$

Among people with characteristics X_i , the treated-to-control ratio is proportional to $p(X_i) / (1 - p(X_i))$.

Weighting controls by these **odds of treatment** makes their covariate distribution resemble the treated distribution, under the population relationship and with a correctly estimated score.

For the ATE, weights are $1 / \hat{p}(X_i)$ for treated observations and $1 / (1 - \hat{p}(X_i))$ for controls. Both groups represent the whole population.

For NSW, we use the **ATT** to estimate the effect on participants.

What Do the Weights Mean?

Control's estimated score	ATT weight
0.20	0.25
0.50	1.00
0.80	4.00
0.95	19.00

A control with a score of 0.80 receives sixteen times the weight of one with a score of 0.20.

The last row links weighting directly to **common support**: a few controls can dominate the weighted comparison.

IPW Regression: Weighted Least Squares

In PP5000, we derived OLS using the **method of moments**. The same estimates minimise the sum of squared residuals:

$$\min_{b_0, b_1} \sum_i (Y_i - b_0 - b_1 D_i)^2 .$$

Every observation receives equal weight. **Weighted least squares** minimises

$$\min_{b_0, b_1} \sum_i w_i (Y_i - b_0 - b_1 D_i)^2 .$$

An observation with a larger weight contributes more to the sum. The minimising values are $\hat{\beta}_0$ and $\hat{\beta}_1$.

For IPW, use the ATT weights to make the controls resemble the treated group.

The Coefficient Is a Difference in Means

With only an intercept and treatment indicator,

$$\hat{\beta}_1 = \underbrace{\frac{1}{N_1} \sum_{i:D_i=1} Y_i}_{\text{treated mean}} - \underbrace{\frac{\sum_{i:D_i=0} w_i Y_i}{\sum_{i:D_i=0} w_i}}_{\text{weighted control mean}} .$$

The intercept estimates the weighted control mean; intercept plus treatment coefficient estimates the treated mean.

Dividing by the sum of control weights makes the second term a **weighted average**. Under our assumptions, it estimates the treated group's mean outcome without treatment.

Check the Comparison Before Interpreting the Effect

1. Plot the propensity scores for both groups: is there common support?
2. Inspect control weights: does the comparison depend on a few people?
3. Compare covariate balance before and after weighting.
4. Report the estimated effect, its units, and uncertainty.

Use heteroskedasticity-consistent standard errors for the outcome regression. Ordinary WLS standard errors are inappropriate for these treatment weights.

Robust WLS errors treat estimated weights as fixed; fuller inference must also account for estimating the propensity score.

Problem Set 1

The Week 2 practice will be **Problem Set 1**.

Compare the experimental benchmark with estimates using **two nonexperimental comparison samples**:

- Unadjusted and covariate-adjusted OLS.
- ATT weighting via WLS: 1975 earnings alone, then add 1974 earnings.
- Common support, weights, balance, and interpretation.

Submit your **Quarto source and rendered PDF**, combining Python, results, and explanation. Due **Friday 2 October, 12 noon**.

[Problem Set 1: instructions and Quarto answer template](#)

References

- MDRC (1980). *Summary and Findings of the National Supported Work Demonstration*. Tables 4-3, 5-3, 6-3, 7-3.
- LaLonde, R. J. (1984). *Evaluating the Econometric Evaluations of Training Programs with Experimental Data*. Princeton IRS Working Paper 183, Chapter 2.
- LaLonde, R. J. (1986). “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *AER* 76(4): 604–620.
- Rosenbaum, P. R., and D. B. Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1): 41–55.
- Dehejia, R. H., and S. Wahba (1999). “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs.” *JASA* 94(448): 1053–1062.
- King, G., and R. Nielsen (2019). “Why Propensity Scores Should Not Be Used for Matching.” *Political Analysis* 27(4): 435–454. [DOI](#).

Textbook readings and links: [Week 2 notes](#).