

Problem Set 1: Matching and Propensity-Score Weighting

PP5001 · Martinmas 2026

[Download the PDF](#) · [Download the Quarto answer template](#) · [Week 2 notes](#)

Due: Friday 2 October 2026, 12 noon (UK time).

This is the assessed Week 2 practice. Submit **your completed Quarto source (.qmd) and rendered PDF** through Moodle. Use your student ID, not your name, in the document. Render the whole document before submitting.

Use the notes and LaLonde/Dehejia–Wahba readings to explain your findings. You will compare estimates from OLS with estimates using propensity-score weights to estimate the average effect on the treated (ATT).

Code clarification — 29 September

If you copied the Python template from the Week 2 notes before 29 September, add `"re74"` to its `columns` list and rerun the code from the sample-construction step. Keep this variable in the data even though it first enters the propensity-score model in Question 5.

In the corrected examples, `sample` always contains both treated and comparison observations. `control_group` contains only the comparison observations for the balance calculations. Estimate the propensity-score model on the combined `sample`.

Download and import the data

Download the **Stata (.dta) files** from [Rajeev Dehejia's data page](#). Save them in a folder named `data` beside your Quarto file. Pandas reads this file format directly.

Use `nsw_dw.dta`, under “NSW Data Files (Dehejia-Wahba Sample)”. It contains **185 treated men and 260 experimental controls**. Do not use `nsw.dta`, which is LaLonde's larger 297/425 sample. The sample change also changes the experimental benchmark. Keep the same 185 treated men throughout, including when adding 1974 earnings.

Choose **two** of the six observational comparison samples and download their Stata files:

Comparison sample	Stata filename	Observations
PSID1	psid_controls.dta	2,490
PSID2	psid_controls2.dta	253
PSID3	psid_controls3.dta	128
CPS1	cps_controls.dta	15,992
CPS2	cps_controls2.dta	2,369
CPS3	cps_controls3.dta	429

PSID2 download correction: the link labelled `psid_controls2.dta` on Dehejia's page currently points to PSID1. Use the correct PSID2 file in his archive. If an older HTTP link fails, use HTTPS on [users.nber.org](https://users.nber.org/~rdehejia/data/) with the same `/~rdehejia/data/` path and filename. Check observation counts after importing.

For example:

```
from pathlib import Path
import pandas as pd
```

```
data_dir = Path("data")
experiment = pd.read_stata(
    data_dir / "nsw_dw.dta", convert_categoricals=False
)
```

Use the `treat` indicator to distinguish treated observations (1) from experimental controls (0). Import your chosen comparison files separately, inspect their columns and sample sizes, and construct each observational sample using the experimental treated observations plus that comparison group. `convert_categoricals=False` preserves numeric indicator values rather than converting value labels into categories.

The analysis variables are `treat`, `age`, `education`, `black`, `hispanic`, `married`, `nodegree`, `re74`, `re75`, and `re78`. `data_id` labels the source; it is not a regression covariate. Earnings are annual amounts in **1982 US dollars**; zero earnings are valid observations.

The smaller PSID/CPS samples are alternative subsets, not independent experiments. The archive notes that CPS2/CPS3 are reconstructions rather than exact original subsets; PSID1 contains 2,490 observations rather than the 2,493 in LaLonde's Table 3. Those qualifications do not prevent you from reproducing the **DW experimental benchmark** with `nsw_dw.dta`.

Regression models and reporting

Let D_i denote treatment and Y_i denote 1978 earnings. The unadjusted outcome regression is $Y_i = \beta_0 + \beta_1 D_i + u_i$.

For OLS with controls and the first logit model, use **age, age squared, education, Black, Hispanic, married, no degree, and 1975 earnings**. Use the same explanatory variables in both models and for both chosen comparison samples. In Question 5, add 1974 earnings to the propensity-score model only, holding the sample and every other term fixed. Use an intercept throughout.

In Python formula notation, the covariates are:

```
covariates = (
    "age + I(age ** 2) + education + black + hispanic "
    "+ married + nodegree + I(re75 / 1000)"
)
```

Dividing 1975 earnings by 1,000 helps numerical scaling; it does not change the model's fitted values. Keep the outcome, 1978 earnings, in dollars.

Use **heteroskedasticity-consistent standard errors**:

```
result = model.fit(cov_type="HC1", use_t=True)
```

This applies to the OLS and WLS outcome regressions. The logit fit estimates the weights. HC1 standard errors need not match standard errors printed in the papers. For WLS, these errors treat estimated weights as fixed and do not incorporate uncertainty from estimating the propensity scores.

Create your own tables and do not rely only on regression output. You can use Claude to create attractive tables that incorporate all of your results. Code and interpretation should appear together. You may reuse code from the notes.

Example prompt for Claude

Create a clearly labelled summary table from these Python regression results for my Quarto document. Include the comparison sample, estimation method, treatment-effect estimate, heteroskedasticity-consistent standard error, 95% confidence interval, and sample size. Label earnings effects in 1982 US dollars and round results to two decimal places. Use the fitted model objects to populate the table. Make it fit the page width when printed to PDF, and explain the code so I can understand and adapt it.

1. The experimental benchmark

Using the NSW treated and experimental control observations:

- Report the sample size and mean 1978 earnings for each group. Calculate the difference in mean earnings. Verify that it reproduces the experimental benchmark reported by Dehejia and Wahba (1999), Table 3, after rounding to whole dollars.
- Estimate an OLS regression of 1978 earnings on an intercept and treatment status. Report the treatment coefficient, heteroskedasticity-consistent standard error, and 95% confidence interval. Explain its relationship to your difference in means.
- Estimate the OLS regression with the controls listed above on the experimental sample. Compare it with the unadjusted estimate.
- Explain why we use this estimate as a benchmark. What qualification does LaLonde's working paper raise about the selected experimental sample?

2. Replace the experimental controls

Choose **two distinct comparison samples** from PSID1, PSID2, PSID3, CPS1, CPS2, and CPS3. State your choices.

For each choice, combine the **same 185 NSW treated observations** with that comparison sample. Do not include the NSW experimental controls in these observational samples, and do not pool the two comparison samples.

- Report sample sizes and mean 1978 earnings by treatment status.
- Estimate unadjusted and OLS with controls using the explanatory variables listed above. Report treatment coefficients, HC1 standard errors, and 95% confidence intervals.
- Compare both estimates with the unadjusted experimental benchmark. Explain what changes when you replace the comparison group and when you add covariates.

Use the same explanatory variables for both samples.

3. Estimate propensity scores and examine common support

For **each** observational sample:

- Using the [logit example in the notes](#), regress treatment status on the explanatory variables listed above. Use `.predict()` to obtain each observation's estimated probability of belonging to the treated group: its propensity score.
- Plot treated and comparison-group propensity scores using the same bins between zero and one. Use separate panels and show density (`density=True`) so the different sample sizes do not obscure the comparison.

- c. Describe common support. Where are treated observations poorly represented by controls? Report how many treated scores lie outside the observed minimum-to-maximum range of control scores. Explain why being inside this range is not sufficient evidence of good overlap.

Keep all observations and use the predicted probabilities as estimated. If Python reports a warning or cannot estimate the model, check your data import and regression code. You can ask Claude to explain the message and help diagnose the problem. Make sure you understand any suggested changes before using them.

4. Weighting to estimate the effect on the treated

For each observational sample:

- a. Give every treated observation weight 1. For each comparison observation, divide its estimated probability by one minus that probability:

$$w_i = \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} \quad \text{for } D_i = 0.$$

Report the median, 95th percentile, and maximum control weight. Relate these to the score distributions.

- b. Estimate weighted least squares of 1978 earnings on **an intercept and treatment status only**, using these weights. Report the coefficient, HC1 standard error, and 95% confidence interval.
- c. Calculate the treated mean minus the weighted control mean directly and verify that it equals the WLS treatment coefficient. Explain why the denominator of the control mean is the sum of control weights.
- d. Produce a balance table for age, education, Black, Hispanic, married, no degree, and 1975 earnings. Show the treated mean, unweighted control mean, weighted control mean, and standardised mean differences before and after weighting. Use the same unweighted pooled standard deviation as the denominator before and after weighting, as in the notes.
- e. Does weighting improve balance? Identify any important remaining discrepancies. Does balance on these variables establish that selection bias has been eliminated?

These HC1 standard errors treat the estimated weights as fixed. Note this approximation when reporting uncertainty.

5. What does pre-treatment earnings history add?

For each of your two observational samples:

- a. Re-estimate the propensity-score model, adding 1974 earnings to the explanatory variables. In formula notation, add `I(re74 / 1000)`. Keep the same observations and all other terms unchanged.
- b. Recalculate the ATT weights and estimate the treatment-only WLS regression again. Report its coefficient, HC1 standard error, and 95% confidence interval alongside the 1975-only result and experimental benchmark.
- c. Repeat the overlap plots and weight summaries. Compare balance for both 1974 and 1975 earnings under the two sets of weights, and check the other covariates too. Use a fixed unweighted denominator for each covariate's standardised mean differences.
- d. Why might conditioning on earnings in both 1974 and 1975 improve the comparison? What otherwise unobserved differences might these variables capture? Does including them guarantee that selection bias has been eliminated?

e. Explain why keeping the sample fixed matters for interpreting the change in estimates.

6. What have we learned?

Combine your results in one table: the experimental sample and your two observational samples, with unadjusted OLS, adjusted OLS, and (for the observational samples) both ATT weighting specifications: with 1975 earnings alone and with 1974 and 1975 earnings. Clearly label estimates and standard errors; include sample sizes.

Write a short comparison addressing:

- Which estimates are closest to the experimental benchmark? Consider magnitudes and uncertainty.
- Does improved covariate balance necessarily bring the estimate closer to the benchmark?
- Which of your two comparison samples provides the more convincing comparison, and why? Use the overlap and balance evidence as well as the estimates.
- What assumptions are needed to interpret the weighted coefficient as the ATT? What could still go wrong?
- What would you tell a policymaker who had only these observational data and no experimental benchmark?

Support your conclusion with evidence from your results.

Before submitting

Render from a fresh session. Check that all tables and figures appear, both comparison samples use the same treated observations, all requested questions are answered, and your source contains everything needed to reproduce the PDF using the original downloaded Stata files.

Rendering your work

Render your Quarto document directly to PDF. Check that tables fit the page width and that all columns are readable.