Experimental Analysis of Public Policies

Session 2

Selection, Regression, Matching, and Inference

Etienne Dagorn - Universite de Lille

Session Objectives

By the end of this session, students should be able to:

- → recall the selection-bias decomposition from Session 1 and apply it to new observational comparisons;
- → interpret a regression coefficient as a conditional comparison;
- → use the omitted-variable-bias formula to predict the direction of bias;
- → explain why post-treatment controls are dangerous;
- → define matching, propensity scores, balance, overlap, and common support;
- → state the conditional-independence and overlap assumptions;
- → choose an appropriate level for clustering and separate inference from identification.

Roadmap

Roadmap

From Session 1 to Session 2

- → Session 1 defined the causal estimand and the missing counterfactual.
- → Session 2 studies non-experimental comparisons using observed covariates.
- → The central question is whether treated and untreated units can be made comparable with data we observe.

Central question

When can controls, regression, or matching turn an observational comparison into a causal comparison?

Why Selection Persists

- → Policies are often targeted to people with high need.
- → Participants may be more motivated, informed, or constrained.
- → Institutions may choose who receives treatment first.
- → These forces affect both treatment and outcomes.

Implication

Observed outcome differences usually combine treatment effects and pre-existing differences.

Empirical habit: write one upward-selection story and one downward-selection story before deciding which is more plausible.

Naive Difference

$$\Delta^{naive} = E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$$

- → This comparison is descriptive.
- → It is causal only if untreated potential outcomes are comparable across groups.
- → Comparability is a design claim, not an automatic result of large samples.

Recall: Selection-Bias Decomposition

$$\Delta^{naive} = ATT + \{E[Y_i(0) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]\}$$

- $\rightarrow$ $ATT = E[Y_i(1) - Y_i(0) \mid D_i = 1]$.
- $\rightarrow$ The second term is the Session 1 selection-bias term.
- $\rightarrow$ Today we ask whether observed covariates X_i can make that term plausibly zero.

Interpretation: selection bias is a statement about $Y_i(0)$, not about observed Y_i.

Direction of Bias

Selection story	Likely bias
More motivated pupils seek tutoring	upward
Weak pupils are targeted for tutoring	downward
Rich schools adopt early	upward for school resources
Badly hit areas receive aid first	downward for recovery outcomes

Reasoning

Always ask whether treated units would have had higher or lower untreated outcomes.

Roadmap

Omitted-Variable Bias Intuition

- → Suppose parental motivation affects both tutoring participation and test scores.
- → If motivation is omitted, the treatment coefficient absorbs part of its effect.
- → The coefficient no longer isolates the treatment contrast.

Key point

Controls matter only when they block a real selection channel and are measured before treatment.

Omitted-Variable Bias Formula

Simple regression with no other controls:

$$Y_i = \alpha + \beta D_i + \delta W_i + \varepsilon_i$$

$$\tilde{\beta} = \beta + \delta \frac{\mathrm{Cov}(D_i, W_i)}{\mathrm{Var}(D_i)}$$

- → W_i is an omitted observed variable in this simple algebra.
- → Bias requires two ingredients: W_i affects Y_i and W_i is correlated with D_i.
- → Controls aim to remove these omitted channels.
- → Unobserved U_i remains a threat.

With controls: use the partial relationship between D_i and the omitted variable after conditioning on included covariates.

OVB Sign Rule: Prediction Drill

Sign rule

Bias has the sign of “effect of omitted variable on outcome” times “correlation with treatment”.

Predict before calculating

Parental motivation raises test scores and makes tutoring participation more likely. If motivation is omitted, which way is the tutoring coefficient biased?

Answer

OVB Sign Rule: Prediction Drill

Sign rule

Bias has the sign of "effect of omitted variable on outcome" times "correlation with treatment".

Predict before calculating

Parental motivation raises test scores and makes tutoring participation more likely. If motivation is omitted, which way is the tutoring coefficient biased?

Answer

Upward: the omitted variable is positively related to both Y_i and D_i.

Regression as Conditional Comparison

$$Y_i = \alpha + \beta D_i + \gamma X_i + u_i$$

→ Regression compares treated and untreated units with similar X_i.

→ β is causal only under conditional comparability.

→ The method is statistical; the identifying assumption is conceptual.

Interpretation: a regression table should be read as a set of comparisons, not as a machine that automatically produces causality.

What Controls Can Do

- → Adjust for observed pre-treatment differences.
- → Improve precision when covariates explain outcome variation.
- → Make comparisons more transparent when selection is on observables.
- → Reveal whether the treatment contrast is sensitive to specification.

Boundary: controls help only for variables that are observed, measured before treatment, and substantively connected to selection.

Bad Controls

Definition

A bad control is a variable affected by treatment, or a variable that opens a misleading comparison path.

- → Controlling for attendance after tutoring may remove part of the treatment effect.
- → Controlling for later income in a training evaluation can create post-treatment bias.
- → Prefer baseline covariates measured before treatment.

Rule of thumb: if the policy could have changed the variable, do not include it as an ordinary control without a causal diagram or clear estimand.

Prediction Controls vs Causal Controls

Goal	Control logic
Prediction	include variables that improve forecast accuracy
Causality	include pre-treatment confounders, avoid mediators
Precision	include baseline predictors of the outcome
Transparency	report why each control set is used

Roadmap

Applied Example: Job-Search Workshop

- → Policy: voluntary job-search workshop.
- → Treated: unemployed workers who attend.
- → Outcome: employment after six months.
- → Baseline variables: age, education, unemployment duration, prior earnings, sector.
- → Remaining concern: motivation and informal networks may be unobserved.

LaLonde-style lesson: observational adjustment is most convincing when we can explain why the adjusted comparison would reproduce a credible benchmark.

Conditional Independence

Selection on observables

$$\{Y_i(1), Y_i(0)\} \perp D_i \mid X_i$$

- → After conditioning on X_i, treatment is as good as random.
- → This is not testable directly because potential outcomes are missing.
- → It is credible only if the main selection channels are observed.

Boundary

Controls and matching cannot solve unobserved selection by themselves.

Matching: Motivation

Idea

Instead of imposing a linear regression adjustment, compare treated units to untreated units that look similar before treatment.

→ Matching makes the comparison group more visible.

→ It is useful when treated and untreated groups have different covariate distributions.

→ It still relies on selection on observables.

Class question: for each treated unit, who is the most credible untreated analogue?

Toy Matching Table

Worker	D_i	**Prior earnings**	**Unemp. duration**	Y_i
Amina	1	low	long	1
Boris	1	high	short	1
Chloe	1	low	short	0
David	0	low	long	0
Elise	0	high	short	1
Farid	0	high	long	0

By hand

Match each treated worker to the closest untreated worker. Which treated worker has weak support?

Exact Matching

- → Compare treated and untreated units with identical values of covariates.
- → Example: same grade, same baseline score bin, same school type.
- → Exact matching is intuitive but becomes difficult with many covariates.

Dimensionality

With many continuous variables, exact matches may not exist.

Covariate Balance

- → Balance means treated and comparison units have similar pre-treatment covariate distributions.
- → It is checked before interpreting outcomes.
- → Balance is a diagnostic, not proof that unobserved confounding is absent.

Question

After matching, do treated and control units look comparable on the variables used to justify the design?

Propensity Score

$$p(X_i) = Pr(D_i = 1 \mid X_i)$$

→ The propensity score summarizes observed selection into treatment.

→ Units with similar $p(X_i)$ have similar predicted treatment probability.

→ Matching on the propensity score can reduce a high-dimensional covariate problem.

Important: the propensity score is a tool to build balance; it is not the causal effect and not the outcome model.

Overlap and Common Support

Overlap

For relevant values of X_i, both treated and untreated units must exist:

$$0 < Pr(D_i = 1 \mid X_i) < 1.$$

- → Without overlap, the comparison requires extrapolation.
- → Poor overlap changes the population for which effects are credible.
- → Trimming can improve credibility but narrows the estimand.

Graph reflex: always inspect the propensity-score distribution separately for treated and untreated units.

Matching Estimands

- → Nearest-neighbor matching often targets the ATT.
- → Weighting can target ATE or ATT depending on weights.
- → Trimming for overlap changes the effective target population.

Reporting

State whether the estimate applies to all units, treated units, or units in common support.

Limits of Matching

- → Matching cannot solve unobserved selection.
- → It can discard observations and change the estimand.
- → It can look credible mechanically while missing the main confounder.
- → Standard errors need care because matching or weighting is part of estimation.

Example: matching on age and education will not remove bias from unobserved motivation if motivation drives both attendance and employment.

A Matching Workflow

1. Define treatment, outcome, estimand, and pre-treatment covariates.
2. Estimate or choose a distance between treated and control units.
3. Match or weight units.
4. Check balance and overlap.
5. Estimate the effect on the matched or weighted sample.
6. Discuss remaining unobserved selection.

Order matters: balance and overlap should be assessed before making the outcome result the centre of the discussion.

Choosing Variables for Matching

- → Include variables that jointly affect treatment and outcome.
- → Use baseline outcomes when available.
- → Avoid variables measured after treatment begins.
- → Avoid including too many weak covariates if they destroy overlap.

Tradeoff

More covariates can reduce confounding but also make comparable matches harder to find.

Matching vs Regression

	Regression	**Matching**
Comparison	conditional mean	similar units
Functional form	explicit model	distance or weights
Diagnostics	residuals, sensitivity	balance, overlap
Risk	model dependence	poor matches, trimming

Both require observed confounders to be sufficient.

Roadmap

Inference Questions

- → What is the sampling or assignment process?
- → At what level does treatment vary?
- → Are outcomes correlated within schools, firms, regions, or years?
- → Is the study powered to detect a policy-relevant effect?

Separation: first ask whether the comparison is credible; then ask how uncertain the resulting estimate is.

Standard Errors

- → Standard errors quantify uncertainty, not bias.
- → Robust standard errors address heteroskedasticity.
- → Clustered standard errors address within-cluster correlation.
- → They do not make a weak design causal.

Exam habit

Do not answer an identification problem with a standard-error correction.

Useful phrase: "The estimate may be precise, but the comparison is not credible."

Clustering

Practical rule

Cluster at the level of assignment, and at any level where residual shocks are correlated in ways relevant to the identifying variation.

- → School-level programme: cluster by school.
- → Regional policy: cluster by region.
- → Panel shocks within unit and year: consider two-way clustering.
- → With few clusters, conventional clustered standard errors can still be unreliable.

Bias, Precision, and External Validity

Issue	Question
Bias	Is the comparison causal?
Precision	How uncertain is the estimate?
External validity	For whom and where does it matter?

Separation

A precise biased estimate is still biased. A credible narrow estimate may still be useful.

Matching Workshop: Design and Debrief

Scenario: a voluntary job-search workshop is offered to unemployed workers.

1. Choose five pre-treatment variables for matching.
2. Name one variable that should not be used because it is post-treatment.
3. Predict where overlap will be weakest.
4. State whether the target estimand is ATE, ATT, or an overlap-population effect.

Diagnostic	**Interpretation**
Balance improves	observed comparison becomes more credible
Overlap is poor	effect is local to comparable units
Large hidden confounder	matching may still be biased

Common Mistakes

Watch for these

- → Matching on variables affected by treatment.
- → Reporting balance only after looking at outcomes.
- → Treating the propensity score model as the causal model.
- → Ignoring poor overlap.
- → Confusing statistical precision with identification.

Practice Bridge

- → Theory sheet: `exercises/theory/session_02_theory.md`.
- → QCM: `assessments/qcm/session_02_qcm.md`.
- → Mini code task: estimate a propensity score, inspect overlap, and compute a simple matched comparison.
- → Final-exam skill: explain whether a coefficient or matched contrast deserves a causal interpretation.

In-class sequence: compute the naive gap, add controls, inspect overlap, then decide whether the causal claim improved.

Roadmap

Selection That Remains

Regression as a Conditional Comparison

Matching, Balance, and Common Support

Inference and Interpretation

Bridge to Randomisation

Design Logic

Question	Can observational comparisons recover a causal effect?
Ideal experiment	Randomise treatment within covariate strata.
Counterfactual	Untreated outcomes for comparable treated units.
Variation	Regression controls, matched units, or weights based on pre-treatment covariates.
Assumption	Conditional independence and overlap.
Threat	Unobserved selection, bad controls, poor common support, wrong inference level.

Reading cue: keep asking which observed variables make selection plausible and which unobserved variables remain outside the design.

Key Takeaways

- → Selection bias is a difference in untreated potential outcomes.
- → Regression and matching are comparison strategies, not designs by themselves.
- → Matching requires conditional independence and overlap.
- → Balance and common support are essential diagnostics.
- → Standard errors address uncertainty, not identification failure.

One-sentence memory aid: observational designs can be useful, but their credibility lives in the selection story.

Transition to Session 3

- → Regression and matching improve comparisons only if observed covariates are enough.
- → Balance and overlap are diagnostics; they do not reveal unobserved selection.
- → Standard errors quantify uncertainty around an estimator, not the credibility of the identifying assumption.

Next

What if we create comparability directly through random assignment?

Bridge: Session 3 asks what randomisation solves, and what implementation problems remain after assignment is clean.