

# Measurement Error and Placebo Tests

**Note on data.** This problem set uses **synthetic** (simulated) data generated in the provided Python tutorial script. The goal is to understand how measurement error distorts inference and how placebo tests help diagnose pipeline artifacts—not to make substantive claims about real-world outcomes.

## Conceptual Questions

Please write three to ten sentence explanations for each of the following questions. **You are only required to answer ONE of the two questions below.**

1. Define **measurement error** and explain why it can bias regression estimates. In your answer, distinguish between (i) measurement error in a *predictor* that is not a confounder and (ii) measurement error in a *confounder*. Explain why the second case can bias the estimated treatment effect even when the model “controls for” the noisy confounder.
2. Placebo tests are often described as “pipeline diagnostics.” Explain what an **outcome placebo** and a **treatment permutation placebo** are. For each, describe what you would expect to see if the analysis pipeline is behaving well, and give one concrete reason a placebo test might fail even when the underlying causal effect is truly zero.

## Applied Exercises

Use the code in the week’s code tutorial and the lecture slides to answer the following questions.

3. **Measurement error simulation: bias and attenuation.** Using the provided tutorial script (`measurement_error_placebo_tests.py`):
  - Run the script end-to-end and confirm it writes the outputs (CSV + figures) to `outputs/` and `figures/`.
  - Using `outputs/measurement_error_results.csv`, make a clean table that reports, for each `sigma_u`:
    - (a) `tau_oracle_mean`, `tau_naive_mean`, `tau_cal_mean`, and
    - (b) `beta_oracle_mean`, `beta_naive_mean`, `beta_cal_mean`.
  - In 6–10 sentences, interpret the two figures:
    - (a) `figures/measurement_error_tau_vs_sigma.png` and
    - (b) `figures/measurement_error_beta_vs_sigma.png`.Your interpretation must address:
    - (a) how and why the naive estimate of the treatment effect changes as `sigma_u` increases,
    - (b) attenuation of the confounder coefficient (why the estimated confounder effect shrinks), and
    - (c) the difference between the “oracle” and “naive” estimands in this simulation.

4. **Validation subsample and regression calibration.** The tutorial implements a simple calibration approach using a validation subsample where `x_true` is observed.
  - Change the validation share to **three** values (e.g., 0.05, 0.20, 0.50) by editing `validation_share`.
  - Hold `sigma_u` fixed at **one** moderate value (e.g., `sigma_u = 1.0`) and rerun the simulation for each validation-share setting (keeping all other settings the same).
  - Report, for each validation share, the mean calibrated treatment estimate `tau_cal_mean` and compare it to the naive estimate `tau_naive_mean`.
  - In 6–10 sentences, explain what changes as the validation sample grows. Your discussion must include:
    - (a) why calibration can help when the confounder is measured with error,
    - (b) why calibration is not “magic” (what assumptions it relies on), and
    - (c) one reason calibration might fail or remain biased in real social data.
5. **Placebo tests: outcome placebo and treatment permutation.** Using the same script:
  - Outcome placebo: report the estimated coefficient on `d` in the placebo regression and explain why it should be close to zero.
  - Treatment permutation placebo: run the permutation test and report:
    - (a) the observed naive estimate `tau_hat_obs`,
    - (b) the empirical two-sided p-value,
    - (c) a histogram figure of the permutation distribution with the observed estimate marked.
  - In 6–10 sentences, interpret what the permutation distribution is telling you. Your interpretation must address:
    - (a) what the null hypothesis is in the permutation test,
    - (b) what it means if the observed estimate is extreme relative to the permutation distribution, and
    - (c) one way placebo tests can detect pipeline problems (e.g., leakage, overfitting, coding errors).
6. **Challenge Question (Optional — if you finish early):** Make the placebo tests more informative by strengthening the null distribution. Choose **ONE** option:
  - (a) **Block permutation.** Permute `d` *within bins* of `x_obs` (e.g., quintiles) instead of permuting globally. Compare the resulting placebo distribution to the global permutation distribution and discuss (5–10 sentences) what changes and why.
  - (b) **P-value calibration.** Repeat the entire simulation under a *true null* where `tau = 0` (set the treatment effect to zero in the DGP), and run the permutation placebo test across at least 100 simulated datasets. Plot the histogram of empirical p-values. In 5–10 sentences, interpret whether the p-values look approximately uniform and what that implies about your testing procedure.