ECON 3500 Econometrics and Applications
Spring 2026

In-Class Activity: Regression Validity — Solutions

Chapter 9 — Assessing Studies Based on Multiple Regression
Time: ~15-20 minutes

Your Job

Each example below is a research study. All six are based on real papers, but the descriptions below are simplified for class use.

For each study:

1. What is the **goal**?
   - Causal inference
   - Forecasting
2. What is the **main problem**?
   - Omitted variable bias
   - Wrong functional form
   - Errors-in-variables bias
   - Sample selection bias
   - Simultaneous causality bias
   - External validity only / not mainly an internal-validity problem
3. Why is that the right diagnosis?
4. What is one concrete fix or improvement?

Quick Diagnosis Guide

If the problem is…	Ask yourself…
OVB	Is there some omitted factor that affects Y and is correlated with X?
Wrong functional form	Did we force a straight-line relationship when the true relationship is curved or interactive?
Measurement error	Is X or Y measured noisily, inaccurately, or systematically wrong?
Sample selection	Are some observations missing because of the outcome or some unobserved factor tied to it?
Simultaneous causality	Does Y also affect X?
External validity	Even if the study is internally valid, would the result generalize to a different setting?

Example 1: Catholic Schooling and Educational Attainment

A researcher studies whether attending a Catholic high school increases graduation and college attendance. Students who attend Catholic schools may also come from families that are more motivated, more religious, or more education-focused to begin with.

- Goal: **Causal inference**
- Diagnosis: **Omitted variable bias**
- Why: Students who attend Catholic schools are selected. Family motivation, religiosity, discipline, and neighborhood context may affect both school choice and later attainment.
- Fixes:
  - Add better controls
  - Use a credible IV or lottery-style design
  - Compare similar students more carefully

Example 2: Oregon Medicaid Lottery

The Oregon Health Insurance Experiment used a lottery to study the effects of Medicaid for low-income uninsured adults in Oregon. A policymaker wants to use those estimates to predict what the effects would be in a very different state with different hospitals, demographics, and eligibility rules.

- Goal: Usually **causal inference** in the original study, but the policymaker's question is about **external validity**
- Diagnosis: **External validity only / not mainly an internal-validity problem**
- Why: The question is whether Oregon lottery estimates transport to a very different setting. Students should talk about hospitals, baseline uninsured rates, take-up, and the local policy environment.
- Fixes:
  - Replicate in more settings
  - Compare institutional context
  - Ask whether the treated and target settings are genuinely comparable

---

Example 3: Survey Earnings vs. Administrative Records

Bound and Krueger compare workers' self-reported earnings in surveys to administrative earnings records. Suppose a researcher estimates the effect of earnings on some outcome using only the self-reported survey measure.

- Goal: **Causal inference** or prediction; either answer is acceptable if justified
- Diagnosis: **Errors-in-variables bias**
- Why: Self-reported earnings differ from administrative records. The observed regressor may contain measurement error, and Bound-Krueger show that it is not purely classical.
- Fixes:
  - Use administrative records
  - Validate survey responses
  - Be cautious about assuming classical attenuation only

Example 4: Wages of Married Women

In Heckman's classic sample-selection setup, wages are only observed for married women who choose to work. A researcher regresses wages on education using only women with observed wages.

- Goal: **Causal inference**
- Diagnosis: **Sample selection bias**
- Why: Wages are only observed for women who work. Selection into employment depends on unobservables that may also affect wages.
- Fixes:
  - Model the selection process
  - Use Heckman-style correction methods
  - Gather information on nonworkers if possible

Example 5: Children and Mothers' Labor Supply

A researcher regresses a mother's labor supply on the number of children she has and finds that women with more children work less. He concludes that having another child reduces labor supply by exactly that amount.

- Goal: **Causal inference**
- Diagnosis: **Simultaneous causality bias**
- Why: Fertility affects labor supply, but labor supply choices may also affect fertility decisions. Family preferences and timing decisions tie the two together.
- Fixes:
  - IV
  - Natural experiment
  - Exogenous variation in family size

Example 6: Earnings and Experience

Following the classic earnings literature, a researcher regresses log earnings on years of schooling and years of labor-market experience. She includes experience only as a linear term, even though the earnings profile appears to rise early in the career and then flatten.

- Goal: **Prediction** or description, though students may argue causal inference if they justify it carefully
- Diagnosis: **Wrong functional form**
- Why: A linear term imposes a constant marginal effect of experience, but the classic earnings profile is concave.
- Fixes:
  - Add experience squared
  - Use logs
  - Plot the data first

Final Checkup

Choose **one** of the six studies above and answer:

If you were the journal referee, would you trust the causal claim? Why or why not?

Answers will vary — the key is applying the correct diagnosis and explaining how it undermines or qualifies the study's conclusions.