Instrumental Variables

Justin S. Eloriaga
ECON 521

Big Picture

- Goal: Estimate the causal effect of receiving a push notification on in-app purchases.
- Challenge: "Push delivered" is *endogenous* (delivery correlates with user income/tech), so naive regressions are biased.
- Solution: Use **Instrumental Variables (IV)** with **assignment to push** as the instrument for **delivery**.
- We compare OLS vs. 2SLS, explain *LATE*, check instrument strength, and illustrate weak-IV pitfalls via simulation.

Setup and Intuition (in Words)

- Variables:
  - `push_assigned`: randomized assignment to get a push.
  - `push_delivered`: whether the push actually reached the user.
  - `in_app_purchase`: outcome (USD spent).
  - `income` (unobserved): affects both delivery (device/OS/network) and purchases.
- Problem: `push_delivered` is more likely among higher-income users; income also raises spending $\Rightarrow$ upward bias if we regress spending on delivery directly.
- Instrument idea: `push_assigned` changes the chance of delivery but, by design, is independent of income and affects spending only through delivery.

Identification Assumptions (IV)

- **Relevance:** $\text{Cov}(Z, D) \neq 0$ (assignment affects delivery).
- **Exclusion:** Z (assignment) affects purchases only via D (delivery), not directly.
- **Independence:** Z is as-good-as random w.r.t. unobservables (e.g., income).
- **Monotonicity (for LATE):** No defiers (nobody who would get delivered if *not* assigned but *not* get delivered if assigned).

Data & Preprocessing

- Data: `app_engagement_push.csv` (from *Causal Inference for the Brave and True* materials).
- Purchases scaled to whole USD (e.g., divide by 10 and round).
- Quick descriptive stats (in the notebook) confirm sensible ranges for delivery, assignment, and spending.

Naive OLS and Why It's Biased

- OLS spec (notebook): `in_app_purchase` $\sim$ `push_assigned` $+$ `push_delivered`.
- Empirical pattern: the coefficient on `push_delivered` is about \$2–\$3 (e.g., $\sim$ \$2.76 in the run shown).
- Interpretation: Biased upward because delivery is more likely for users with better devices/income, who also spend more.
- Takeaway: OLS conflates the causal effect of delivery with selection on unobservables.

2SLS Structure (What the Notebook Does)

- **Stage 1:** Predict delivery from assignment

$$D_i = \pi_0 + \pi_1 Z_i + v_i, \quad Z = \texttt{push\_assigned},\ D = \texttt{push\_delivered}.$$

- **Stage 2:** Regress purchases on *predicted* delivery

$$Y_i = \alpha + \beta \widehat{D}_i + u_i, \quad Y = \texttt{in\_app\_purchase}.$$

- Compact formula: $\beta_{\mathrm{IV}} = \dfrac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)}$ (Wald/IV estimator).
- In the notebook this is done via `pyfixest` IV syntax.

2SLS Results and Interpretation

- The 2SLS estimate is *much smaller* than OLS (about \$0.30 vs. \$2–\$3).
- This aligns with the bias story: OLS overstated the effect by mixing in income/device advantages.
- **What 2SLS estimates: LATE.** It's the effect for *compliers*: users whose delivery status is changed by assignment.
- In this context, compliers are plausibly users on the margin of delivery (e.g., with devices/settings where assignment matters); they may skew wealthier than never-takers, which can shape the LATE.

First-Stage Strength (Notebook Diagnostics)

- The notebook reports very large first-stage F-statistics in the real data (traditional $\approx 12{,}846$, robust $\approx 12{,}557$), indicating a **very strong** instrument in that run.
- Rule of thumb: first-stage F > 10 suggests weak-IV concerns are limited (formal Stock–Yogo tests are preferable when available).
- The robust F also accounts for heteroskedasticity/serial correlation in errors.

Why LATE (Conceptual Slide)

- Compliance types under monotonicity:
  - Always-takers: delivered regardless of assignment.
  - Never-takers: never delivered regardless of assignment.
  - Compliers: delivered *if* assigned, not delivered otherwise.
- 2SLS identifies $E[Y(1) - Y(0) \mid \text{complier}]$.
- Policy relevance: LATE targets the group whose behavior changes when assignment changes (e.g., users reachable specifically because they were assigned a push).

Weak-IV Simulation (What the Notebook Shows)

- Synthetic design: $D = \beta Z + U$, $Y = D + U$, true effect $= 1$, $n = 100$.
- **Weak instrument:** set $\beta = 0.1$ so Z barely moves D.
- Result: the sampling distribution of the IV estimator is heavy-tailed and far from normal.
- Size distortion: using a normal approximation yields a rejection frequency $\approx 23\%$ at nominal 5%—dramatically over-rejecting.
- Lesson: with weak instruments, 2SLS can be unreliable; use stronger instruments or weak-IV robust methods (e.g., Anderson–Rubin, *LIML*, conditional likelihood ratio).

Practical Checklist (Reflecting the Notebook)

- **Causal story:** Make a clear DAG in words; defend exclusion.
- **First stage:** Report $\widehat{\pi}_1$ and F-stats (traditional and robust).
- **2SLS reporting:** Show $\widehat{\beta}_{\text{2SLS}}$, CIs, and interpret as LATE.
- **Sensitivity:** Consider heteroskedasticity-robust SEs; explore robustness to controls (if allowed by design).
- **Weak-IV guardrails:** Watch F-stats; if low, pivot to weak-IV robust inference or better instruments.

Key Takeaways

- OLS overstated the causal effect of delivery on spending due to selection on unobservables.
- IV using randomized assignment corrects this and targets the causal effect for compliers (LATE), yielding a much smaller and more credible estimate.
- Instrument strength matters: strong in the real data run; the simulation warns how weak IVs break standard inference.

Appendix: Notation & Formulas

- Wald/IV: $\beta_{\mathrm{IV}} = \dfrac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)}$.
- Two-Stage Least Squares:

$$\text{Stage 1: } D_i = \pi_0 + \pi_1 Z_i + v_i, \quad \text{Stage 2: } Y_i = \alpha + \beta \widehat{D}_i + u_i.$$

- LATE (Imbens–Angrist): effect for compliers under independence, exclusion, and monotonicity.

Appendix: What Each Notebook Block Does

- **Imports/Config:** `numpy`, `pandas`, `matplotlib`, `networkx`, `pyfixest`; high-res plotting.
- **Causal story (text):** Explains why delivery is endogenous (income/tech).
- **Data load:** Read `app_engagement_push.csv`; scale purchases to USD.
- **OLS run:** Demonstrates upward-biased coefficient on `push_delivered`.
- **2SLS run:** `push_delivered` instrumented by `push_assigned`; presents first-stage and second-stage results.
- **Diagnostics:** Traditional and robust first-stage F statistics.
- **Weak-IV sim:** Sets small β so Z weakly predicts D; shows size distortion.

Sources

- Matheus Facure, *Causal Inference for the Brave and True* (Non-Compliance & LATE module) — dataset & example framing.
- Classic weak-IV references: Staiger & Stock (1997); Stock & Yogo (2005).