Lecture 23: Review of the Course

Jeremy Majerovitz (University of Notre Dame)

Econometrics, Fall 2025

Plan for Today

Today is our last day of class!

Review of what we've learned

- Also, try to step back and draw connections
- The methods we have seen are all cousins
- Easier to remember assumptions/properties when you see the family resemblance

Final exam will be cumulative: unlike quizzes, anything is fair game

- Today will give you a head start
- Highlight most important ideas
- But, not a substitute for going through the individual lecture notes
- Cannot possibly cover all material in 75 minutes!

Lecture 1: Review of Statistical Concepts

In principle, this is review material, so exam will not focus on it.

But, a few things are very useful, which we used over and over:

Let X, Y, Z be random variables, and a, b, c be real numbers

Linearity Rules for Expectation and Covariance

- $\mathbb{E}\left[aX + bY\right] = a \cdot \mathbb{E}\left[X\right] + b \cdot \mathbb{E}\left[Y\right]$
- $\text{Cov}\left(X, aY + bZ\right) = a \cdot \text{Cov}\left(X, Y\right) + b \cdot \text{Cov}\left(X, Z\right)$

Rules for Variance and Covariance

- $\text{Cov}\left(X, a\right) = 0$, $\text{Cov}\left(X, Y\right) = \text{Cov}\left(Y, X\right)$, $\text{Var}\left(X\right) = \text{Cov}\left(X, X\right)$
- $\implies \text{Var}\left(a + bX\right) = b^2 \cdot \text{Var}\left(X\right)$
- $\text{Var}\left(X + Y\right) = \text{Var}\left(X\right) + \text{Var}\left(Y\right) + \text{Cov}\left(X, Y\right)$

Lecture 1: Review of Statistical Concepts

Law of Iterated Expectations: $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$

- If I know average height by gender, and I know the probability of each gender, then I can compute average height
- Can use for other averages, like average treatment effects!

Law of Total Variance

- $\text{Var}(Y) = \text{Var}(\mathbb{E}[Y \mid X]) + \mathbb{E}[\text{Var}(Y \mid X)]$
- Useful for understanding the R^2 (but less important overall for this course)

Important Distinction: Sample vs. Population

- Statistics is about inferring population parameters from sample of data

Lecture 2: Asymptotics

What happens as our data set grows large?

Do not worry about the technical definitions or material, just big ideas:

- Convergence in Probability: As sample grows large, estimator concentrates around a point
- Convergence in Distribution: As sample grows large, distribution of estimator (cdf) converges to particular distribution
- IID Sampling: Independent and Identically Distributed
- Law of Large Numbers: With i.i.d. data, sample mean converges in probability to population mean:$\bar{X}_n \xrightarrow{p} \mathbb{E}[X_i]$
    - Sample mean is "consistent"
- Central Limit Theorem: With i.i.d. data, sample mean converges to normal distribution:$\sqrt{N}\left(\bar{X}_n - \mathbb{E}[X_i]\right) \xrightarrow{d} N\left(0, \text{Var}\left(X_i\right)\right)$
    - Sample mean is "asymptotically normal"

Lecture 3: Ordinary Least Squares

$$Y_i = \alpha + \beta_1 X_{1,i} + \cdots + \beta_K X_{K,i} + \varepsilon_i$$

Ordinary Least Squares solves:

$$\min_{\alpha,\beta} \frac{1}{N} \sum_{i=1}^{N} (Y_i - \alpha - \beta_1 X_{1,i} - \cdots - \beta_K X_{K,i})^2$$

Solution $(\hat{\alpha}, \hat{\beta})$ solves the (in-sample) orthogonality conditions:

$$\frac{1}{N} \cdot \sum_{i=1}^{N} \hat{\varepsilon}_i = 0, \quad \text{and} \quad \widehat{\text{Cov}}\,(\hat{\varepsilon}_i, X_{k,i}) \qquad\qquad = 0 \text{ for all } k = 1, ..., K$$

With single regressor: $\hat{\beta} = \frac{\widehat{\text{Cov}}(X_i, Y_i)}{\widehat{\text{Var}}(X_i)}$

Population counterpart is the Best Linear Predictor:

$$\min_{\alpha,\beta} \mathbb{E}\left[(Y_i - \alpha - \beta_1 X_{1,i} - \cdots - \beta_K X_{K,i})^2\right]$$

Lecture 4: Conditions for Unbiasedness and Consistency of OLS

Three Conditions for Consistency $\left(\hat{\beta} \xrightarrow{p} \beta\right)$:

- No perfect collinearity
- IID sampling
- Orthogonality condition
    - $\mathbb{E}\left[\varepsilon_i\right] = 0$ and $\text{Cov}\left(X_{k,i}, \varepsilon_i = 0\right)$ for all $k = 1, ..., K$

For unbiasedness $\left(\mathbb{E}\left[\hat{\beta}\right] = \beta\right)$, slightly stronger condition replaces orthogonality:

- Zero conditional mean: $\mathbb{E}\left[\varepsilon_i \mid X_{1,i}, ..., X_{K,i}\right] = 0$
- We focus on consistency in this course: don't need to worry about this detail for the exam!

Lecture 4: Conditions for Unbiasedness and Consistency of OLS

Why is OLS consistent?

- No perfect collinearity ensures unique solution to OLS (unique $\hat{\beta}$ exists)
- IID sampling ensures OLS converges to best linear predictor $\left(\hat{\beta} \xrightarrow{p} \beta^{BLP}\right)$
- Orthogonality condition ensures that best linear predictor is the causal effect we are interested in $\left(\beta^{BLP} = \beta\right)$

Plug-in Principle: To construct a good estimator, replace population objects with their sample analogues

We love sample means: When the estimator is a (continuous) function of sample means, it will be consistent and asymptotically normal

This is reason behind why all our estimators work!

Lecture 5: Omitted Variable Bias

Suppose that the true causal model is:

$$Y_i = \alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

But instead, econometrician uses OLS to naively estimate:

$$Y_i = \tilde{\alpha} + \tilde{\beta}_1 X_{1,i} + \tilde{\varepsilon}_i$$

Even in large sample, estimator is biased:

$$\tilde{\beta}_1 = \frac{\text{Cov}\left(X_{1,i}, Y_i\right)}{\text{Var}\left(X_{1,i}\right)} = \beta_1 + \beta_2 \cdot \underbrace{\frac{\text{Cov}\left(X_{1,i}, X_{2,i}\right)}{\text{Var}\left(X_{1,i}\right)}}_{\text{Regression of } X_{2,i} \text{ on } X_{1,i}}$$

Bias from omitted variables depends on:

- Causal effect of omitted variable on outcome, β_2
- Covariance between omitted variable and regressor
- Use formula to determine the sign and/or if it is zero
- Memorize this formula

Lecture 6: More on Correlation vs. Causation

Other things that cause violations of orthogonality condition:

- Measurement Error
  - Classical measurement error in regressor leads to attenuation (bias towards zero)
  - Classical measurement error in outcome is not a problem
- Reverse Causality/Simultaneity
  - Want to measure effect of X on Y, but pick up effect of Y on X
  - Big problem for trying to distinguish supply and demand!
- Selected Sample
  - E.g. If only observe when $Y_i > c$, then OLS will be biased

Don't need to memorize these formulas, just main idea

Lecture 7: Interpreting Regressions and Regression Tables

Sometimes we take (natural) logs

- A 0.01 change in log variable is 1 log point $\approx 1\%$

Sometimes we make dummy variables

- Equal to 1 if something is true
- Dummy Variable Trap: Must have an omitted category to avoid perfect collinearity
- Interpret coefficient relative to omitted category

Also can have interaction effects

- Coefficient on $\text{Female}_i \times \text{Schooling}_i$ is effect of schooling for women minus effect of schooling for men
- Always include the main effects too

Please also take a look at how to read regression tables

- Common place to lose points on the quizzes!

Lecture 8: Asymptotic Distribution of OLS

Same assumptions as for consistency:

- No perfect collinearity, IID sampling, and orthogonality condition

OLS is asymptotically normal:

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, V\right)$$

$$V = \frac{\mathrm{Var}\left(\left(X_i - \mathbb{E}\left[X\right]\right)\varepsilon_i\right)}{\mathrm{Var}\left(X_i\right)^2}$$

You don't need to memorize this formula, but useful to remember:

- Variance of OLS is higher when variance of ε_i is high
- Variance of OLS is lower when variance of X_i is high

Lecture 9: Heteroskedasticity and Robust Standard Errors

Homoskedasticity vs. Heteroskedasticity:

- Homoskedasticity is when variance of ε_i does not depend on X_i
- Heteroskedasticity is when error is not homoskedastic

Partly for historical reasons, "default" SE only valid under homoskedasticity

- Robust estimator is consistent in either case!
- Always use "`, rob`" in Stata

Robust estimator just uses plug-in principle:

$$\hat{V}_{\text{Robust}} = \frac{\widehat{\text{Var}}\left(\left(X_i - \bar{X}\right)\hat{\varepsilon}_i\right)}{\widehat{\text{Var}}\left(X_i\right)^2}$$

$$\widehat{\text{SE}}_{\text{Robust}}\left(\hat{\beta}\right) = \sqrt{\frac{\hat{V}_{\text{Robust}}}{N}}$$

Lecture 10: Hypothesis Tests and Confidence Intervals

Hypothesis testing ingredients:

- Null hypothesis, $H_0 : \beta = \beta_0$
- Test statistic, $\tau = |t|$, $t = \frac{\hat{\beta} - \beta_0}{\widehat{\mathsf{SE}}(\hat{\beta})}$,
- Critical Value, c

Reject the null if $\tau > c$, otherwise "fail to reject"

Asymptotic Normality of OLS + Consistency of SE

- $\implies$ t-statistic is asymptotically $N(0, 1)$ *under the null*
- Test statistic is $|t|$, so need *97.5th* percentile of standard normal for a 5% test

Confidence Interval $= \hat{\beta} \pm c \cdot \widehat{\mathsf{SE}}\left(\hat{\beta}\right)$

Lecture 11: Testing Single Equations

What if we want to test linear equations?

- E.g. $\beta_1 = 2 \cdot \beta_2$ or $\beta_1 + \beta_2 = 10$

Linear combinations of OLS coefficients are normal

- Create linear combination θ so that hypothesis is $\theta = \theta_0$
  - $\beta_1 = 2 \cdot \beta_2 \implies \theta = \beta_1 - 2 \cdot \beta_2$, $\theta_0 = 0$
  - $\beta_1 + \beta_2 = 10 \implies \theta = \beta_1 + \beta_2$, $\theta_0 = 10$
- Compute $\widehat{\text{SE}}\left(\hat{\theta}\right) = \sqrt{\widehat{\text{Var}}\left(\hat{\theta}\right)}$ using variance rules
  - Remember you need the covariances!
- Hypothesis Test: Reject if $\frac{\left|\hat{\theta} - \theta_0\right|}{\widehat{\text{SE}}\left(\hat{\theta}\right)} > c$
- Confidence Interval: $\hat{\theta} \pm c \cdot \widehat{\text{SE}}\left(\hat{\theta}\right)$

Lecture 12: Testing Multiple Equations

Sometimes we want to test multiple equations

- A Rupee is a Rupee in de Mel et al.:

$$\beta_{\text{Small,Cash}} = \beta_{\text{Small,In-Kind}}$$
$$\beta_{\text{Large,Cash}} = \beta_{\text{Large,In-Kind}}$$
$$\beta_{\text{Small,Cash}} = \frac{1}{2}\beta_{\text{Large,Cash}}$$

We cannot just test each of these individually

- If I run three tests at the 5% level, the probability of rejecting at least one is greater than 5%

Instead, run a single test of the joint hypothesis: F-test

- In my view, not useful to memorize the formula
- In practice, always use robust F-test

Lecture 13: Introduction to Instrumental Variables

Instead of assuming $\text{Cov}(X_i, \varepsilon_i) = 0$, find an instrument, Z_i

For single regressor, single instrument:

$$\hat{\beta}^{IV} = \frac{\widehat{\text{Cov}}(Z_i, Y_i)}{\widehat{\text{Cov}}(Z_i, X_i)} = \frac{\text{Reduced Form}}{\text{First Stage}}$$

For consistency and asymptotic normality, we need:

- Relevance: $\text{Cov}(Z_i, X_i) \neq 0$
- IID Sampling
- Orthogonality: $\mathbb{E}[\varepsilon_i] = 0$ and $\text{Cov}(Z_i, \varepsilon_i) = 0$

A valid IV solves OVB, measurement error, and simultaneity

Lecture 13: Introduction to Instrumental Variables

Can break down IV orthogonality into two distinct ideas:

- Exclusion Restriction: Z_i only affects Y_i through X_i
- As-Good-As-Random Assignment: Z_i is assigned in a way that does not create correlation with ε_i

Stepping back, OLS and IV are closely related

- "Relevance" replaces "No Perfect Collinearity"
    - Need to make sure a solution exists
- "Orthogonality Condition" even keeps the same name
    - This is always the key, untestable assumption in econometrics
- Really, OLS is special case of IV, where $X_i = Z_i$

Lecture 14: Two-Stage Least Squares (2SLS)

To implement IV with multiple regressors and instruments, use 2SLS

- In first stage, regress on instruments and controls to construct predicted $\hat{X}_{1,i}, ..., \hat{X}_{K,i}$
- Then, run regression with predicted $\hat{X}_{1,i}, ..., \hat{X}_{K,i}$, and controls
- Just get the correct standard errors from Stata

Need to have at least as many instruments as endogenous regressors:

- Underidentified: Fewer instruments than endogenous regressors (2SLS fails)
- Just identified: As many instruments as endogenous regressors (2SLS works)
- Overidentified: More instruments than endogenous regressors (2SLS works)
  - Overidentification test: Checks if instruments are mutually consistent

Lecture 15: IV Standard Errors and Weak IV

SEs are similar to OLS formula, with tweaks

$$V = \frac{\text{Var}\left(\left(Z_i - \mathbb{E}\left[Z\right]\right)\varepsilon_i\right)}{\text{Cov}\left(Z_i, X_i\right)^2}$$

Plug-in principle for robust SE, testing and CIs all same (uses asymptotic normality)

OLS SE almost always smaller than IV SE

- IV throws out variation in X_i, OLS exploits it all
- But, sometimes IV is necessary for consistency!

If first stage, π, is small relative to $\text{SE}\left(\hat{\pi}\right)$, asymptotics break down

- Distribution heavy-tailed (non-normal), and biased towards OLS

To protect against weak instruments:

- Test first stage (rule-of-thumb is you want $F > 10$)
- Can use weak-instrument robust confidence intervals (not on the exam)

Lecture 16: Potential Outcomes and RCTs

Framework to think more carefully about causality and treatment effect heterogeneity

Treatment Status: $D_i = 0, 1$; Potential Outcomes: $Y_i(0)$ and $Y_i(1)$

Fundamental Problem of Causal Inference

- Only observe $Y_i(D_i)$, don't observe counterfactual

Average Treatment Effect (ATE) $= \mathbb{E}[Y_i(1) - Y_i(0)]$

Average Treatment Effect on the Treated (ATT) $= \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$

$$\begin{aligned}\underbrace{\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]}_{\text{OLS}} = &\underbrace{\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]}_{\text{ATT}} \\ &+ \underbrace{\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]}_{\text{Selection Bias}}\end{aligned}$$

Under random assignment: Selection Bias $= 0$ and ATT $=$ ATE

Lecture 17: The Local Average Treatment Effect

Extend framework to IV, now D_i is function of Z_i

Four types of people: Never Takers, Always Takers, Compliers, and Defiers

Assumptions:

- IID Sampling, Relevance, As-Good-As-Random Assignment, Exclusion Restriction
- Monotonicity: No Defiers

Under these assumptions, IV estimates the Local Average Treatment Effect

- LATE $= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid \text{Complier}\right]$

For IV, need to think carefully: Who are the compliers? + Is LATE the TE we care about?

Lecture 18: Regression Discontinuity

Sharp regression discontinuity: treatment turns on when $X_i \geq c$

$$Y_i = \alpha + \beta_1 \cdot 1\left(X_i \geq c\right) + \beta_2 \cdot \left(X_i - c\right) + \beta_3 \cdot 1\left(X_i \geq c\right) \times \left(X_i - c\right) + \varepsilon_i$$

The treatment effect is β_1

Fuzzy RD instead uses $1\left(X_i \geq c\right)$ as instrument for treatment:

$$Y_i = \alpha + \beta_1 \cdot D_i + \beta_2 \cdot \left(X_i - c\right) + \beta_3 \cdot 1\left(X_i \geq c\right) \times \left(X_i - c\right) + \varepsilon_i$$

Restrict to a narrow bandwidth around the cutoff

Lecture 18: Regression Discontinuity

Assumptions are same as always, but adapted to this case:

- IID Sampling
- As-Good-As-Random Assignment: *Potential* outcomes continuous around cutoff
- Exclusion Restriction: Cutoff only affects outcome through effect on treatment

Extra Assumptions for Fuzzy:

- Relevance: Discontinuous change in treatment probability at cutoff
- Monotonicity: No Defiers

Whose treatment effect do we identify?

- Sharp: ATE at the cutoff $\mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c\right]$
- Fuzzy: LATE at the cutoff $\mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c, \text{Complier}\right]$

Lecture 19: Difference-in-Differences

Treated group gets the treatment, but only in post period

Two differences are better than one:

- Pre/post difference for treated units
- Minus pre/post difference for control units

$$Y_{it} = \alpha + \beta_1 \text{Treated}_i + \beta_2 \cdot \text{Post}_t + \beta_3 \cdot \text{Treated}_i \times \text{Post}_t + \varepsilon_i$$

Identifying Assumptions (Orthogonality):

- Parallel Trends: Potential outcomes under no treatment are on parallel trends
- No anticipation: Treatment does not affect outcomes in pre-period

Under assumptions, diff-in-diff identifies ATT

Parallel trends is untestable, but very informative to check pre-trends

Lecture 20: Fixed Effects

Individual fixed effects estimates separate intercept for each individual

$$Y_{it} = \alpha_i + \beta \cdot X_{it} + \varepsilon_{it}$$

Implement by including dummy variables for each individual (omitting one)

Equivalent to subtracting off each individual's mean:

$$Y_{it} - \bar{Y}_i = \beta \cdot \left(X_{it} - \bar{X}_i\right) + \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)$$

Including individual fixed effects means only using within-individual variation

Can also do time fixed effects (only use within-time variation) or two-way fixed effects

Lecture 21: Clustered Standard Errors

Traditional robust standard error treats each draw as iid

- Assumes $\text{Cov}\left(\left(X_i - \mathbb{E}\left[X\right]\right)\varepsilon_i, \left(X_j - \mathbb{E}\left[X\right]\right)\varepsilon_j\right) = 0$ if $i \neq j$

Clustering: Allows this covariance to be non-zero within cluster, but must be zero across clusters

When do we know that the covariance is zero across clusters?

1. Independent draws of X_i and ε_i across clusters
2. Independent draws of ε_i across clusters
3. Independent draws of X_i across clusters

If none of the above work, probably need to cluster at a higher level

Rule of thumb: Cluster at level of treatment assignment

Key Ideas

Create estimators with the plug-in-principle

- Take what we want to measure in population, and replace with sample analogue

Estimators are functions of sample means $\implies$ consistent and asymptotically normal

- Use asymptotic normality to do CIs and all sorts of hypothesis tests

Key identifying assumption is always some version of orthogonality

- Need to decide what variation in X_i we want to use
- Untestable: Must bring in our knowledge of world to assess plausibility

Conclusion

One week from tomorrow, I hope you still remember the OVB formula

Long term, even if you forget formulas, I hope statistics feels like less of a black box

I hope that this class has:

- Prepared you for elective classes in the major
- Equipped you to use statistics in the future
- And/or to be more sophisticated consumers of statistical claims

Thanks for a great semester, and good luck on the final!