ECON 120B: Synthesized Lecture Notes
Causal Inference and Regression Theory

Compiled from ECON 120B lecture sequence

Contents

1 Scope and Structure **5**

2 Causal Framework and Potential Outcomes **5**
2.1 Prediction vs Causation 5
2.2 Potential Outcomes Setup 5
2.3 Random Assignment and Identification 5
2.4 Regression Representation of Potential Outcomes 6

3 Simple Linear Regression: Setup and OLS Derivation **6**
3.1 Population Model 6
3.2 Least Squares Objective 6
3.3 Orthogonality Geometry 6

4 Finite-Sample Properties Under Exogeneity **7**
4.1 Unbiasedness 7
4.2 Conditional Variance Under Homoskedasticity 7

5 Asymptotic Theory for Simple OLS **7**
5.1 Consistency 7
5.2 Asymptotic Normality 7
5.3 Homoskedastic Special Case 8

6 Heteroskedasticity and Robust Standard Errors **8**
6.1 Why Homoskedasticity is Optional for Consistency 8
6.2 Robust (Eicker-White) Variance in Simple Regression 8

7 Goodness of Fit and Its Limits **8**
7.1 Adjusted R-Squared 8
7.2 Adding Regressors and Mechanical Fit Improvement 9
7.3 Partial R-Squared for a Block of Variables 9

8 Inference for a Single Coefficient **9**
8.1 t Statistic 9
8.2 One-sided Alternatives 9
8.3 Confidence Interval Duality 9
8.4 Exact Finite-Sample t Inference Under Normal Errors 10

9 Special Cases from 120B Additional Topics **10**
9.1 Binary Regressor as Difference in Means . . . 10
9.2 Units of Measurement . . . 10
9.3 Log Specifications . . . 10
9.4 Classical Measurement Error in Regressor . . . 10
9.5 Classical Measurement Error in Dependent Variable . . . 11

10 Multiple Regression Foundations **11**
10.1 Model . . . 11
10.2 Matrix Notation . . . 11
10.3 Interpretation of Partial Effects . . . 11

11 Projection Algebra and Partialling Out **11**
11.1 Hat Matrix and Residual Maker . . . 11
11.2 Frisch-Waugh-Lovell (FWL) Theorem . . . 12
11.3 Interpretation Consequence . . . 12

12 Omitted Variable Bias **12**
12.1 Two-Regressor OVB Formula . . . 12
12.2 Sign Logic . . . 13

13 Finite-Sample and Asymptotic Properties in Multiple Regression **13**
13.1 Conditional Unbiasedness . . . 13
13.2 Asymptotic Distribution . . . 13
13.3 Robust Covariance Estimator . . . 13
13.4 Homoskedastic Covariance in Matrix Form . . . 13
13.5 Gauss-Markov Result (BLUE Statement) . . . 14
13.6 Perfect and Imperfect Multicollinearity . . . 14

14 Functional Form in Multiple Regression **14**
14.1 Polynomials . . . 14
14.2 Continuous-Continuous Interaction . . . 14
14.3 Binary-Continuous Interaction . . . 15
14.4 Categorical Regressors . . . 15
14.5 Exact vs Approximate Log Interpretation . . . 15
14.6 Turning Point in Quadratic Models . . . 15
14.7 Centering in Interaction Models . . . 15

15 Testing Linear Restrictions **15**
15.1 Single Linear Combination . . . 15
15.2 Variance of Linear Combination . . . 16

16 Joint Testing: Bonferroni and F/Wald **16**
16.1 Bonferroni Approach . . . 16
16.2 F/Wald Approach . . . 16
16.3 Overall Significance . . . 16
16.4 Bonferroni Critical Values in Practice . . . 16
16.5 Bonferroni Joint Confidence Intervals . . . 17

17 Worked Derivations and Examples **17**
17.1 Example A: Difference-in-Means via OLS . . . 17
17.2 Example B: OVB Sign . . . 17
17.3 Example C: Linear Restriction Test . . . 17
17.4 Example D: Polynomial Marginal Effect . . . 17
17.5 Example E: Bonferroni with Two Restrictions . . . 18
17.6 Example F: Bonferroni at 1% with Two Restrictions . . . 18
17.7 Example G: Dummy Variable Trap . . . 18
17.8 Example H: Scaling Transformations . . . 18
17.9 Example I: Testing a Nonlinear-and-Interaction Block . . . 18
17.10 Example J: Measurement Error Contrast . . . 19

18 Extended Derivation Workbook **19**
18.1 Finite-Sample OVB Decomposition . . . 19
18.2 One Restriction: Wald, F, and t Equivalence . . . 19
18.3 Linear Combination Confidence Interval . . . 20
18.4 Delta Method for Log-Level Interpretation . . . 20
18.5 Joint Hypotheses in Polynomial-Interaction Models . . . 20
18.6 Interpreting Non-Rejection Correctly . . . 20

19 Assumption and Interpretation Checklist **21**
19.1 Checklist . . . 21
19.2 Common Failure Modes . . . 21

20 Proof Appendix **21**
20.1 Frisch-Waugh-Lovell (Two Regressors, Scalar Form) . . . 21
20.2 Union-Bound Proof for Multiple Testing Control . . . 21
20.3 Sandwich Variance Intuition . . . 22
20.4 R-Squared Equals Squared Correlation in Simple Regression . . . 22
20.5 F Statistic from Restricted and Unrestricted SSR . . . 22
20.6 Consistency Under Classical Error in Y . . . 22

21 Practice Set with Solution Sketches **23**
21.1 Q1: Causal Identification . . . 23
21.2 Q2: OLS Derivation . . . 23
21.3 Q3: OVB Mechanics . . . 23
21.4 Q4: Robust vs Conventional SE . . . 23
21.5 Q5: Joint Restriction . . . 23
21.6 Q6: Interaction Interpretation . . . 23
21.7 Q7: Multiple-Testing Logic . . . 23
21.8 Q8: Elasticity Interpretation . . . 23
21.9 Q9: Measurement Error . . . 23
21.10 Q10: Causal Language Discipline . . . 24
21.11 Q11: Adjusted R-Squared . . . 24
21.12 Q12: FWL Computation . . . 24
21.13 Q13: Multicollinearity . . . 24
21.14 Q14: Dummy Baseline . . . 24
21.15 Q15: Bonferroni 1% Rule . . . 24

21.16Q16: Measurement Error 24
22 Derivation and Proof Summary **24**
23 Bridge to 120C **25**

1 Scope and Structure

These notes synthesize the core ECON 120B sequence:

1. causal framework and potential outcomes,
2. simple OLS derivation and inferential theory,
3. omitted variable bias and multiple regression,
4. interactions, nonlinear terms, and categorical regressors,
5. testing linear restrictions with Bonferroni and F/Wald methods.

The style is derivation-first: each claim is tied to assumptions and stated with the minimal level of proof needed for reliable empirical use.

2 Causal Framework and Potential Outcomes

2.1 Prediction vs Causation

In empirical economics, association can be strong without causal interpretation. A causal claim requires an intervention interpretation: changing X should change Y, holding other channels fixed.

2.2 Potential Outcomes Setup

Let binary treatment $X \in \{0, 1\}$ and potential outcomes $(Y(1), Y(0))$. Observed outcome satisfies:

$$Y = XY(1) + (1 - X)Y(0) = Y(0) + X\big(Y(1) - Y(0)\big).$$

Definition 2.1 (Causal estimands)**.**

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0)], \qquad \text{ATT} = \mathbb{E}[Y(1) - Y(0) \mid X = 1].$$

Remark 2.2. The fundamental problem of causal inference is that for a given unit, only one of $Y(1), Y(0)$ is observed.

2.3 Random Assignment and Identification

Suppose treatment is randomly assigned:

$$X \perp (Y(1), Y(0)).$$

Then

$$\mathbb{E}[Y \mid X = 1] = \mathbb{E}[Y(1)], \qquad \mathbb{E}[Y \mid X = 0] = \mathbb{E}[Y(0)],$$

so

$$\mathbb{E}[Y \mid X = 1] - \mathbb{E}[Y \mid X = 0] = \mathbb{E}[Y(1) - Y(0)] = \text{ATE}.$$

Proof. Under independence, conditioning on X does not change the distribution of potential outcomes. Substitute into observed-outcome identity. $\square$

2.4 Regression Representation of Potential Outcomes

Decompose

$$Y(1) = \mathbb{E}[Y(1)] + u(1), \qquad Y(0) = \mathbb{E}[Y(0)] + u(0).$$

Then

$$Y = \beta_0 + \beta_1 X + u,$$

where

$$\beta_0 = \mathbb{E}[Y(0)], \qquad \beta_1 = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \text{ATE},$$

$$u = u(0) + X\big(u(1) - u(0)\big).$$

If assignment is random, then $\mathbb{E}[u \mid X] = 0$.

3 Simple Linear Regression: Setup and OLS Derivation

3.1 Population Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n.$$

Core identifying condition for causal interpretation:

$$\mathbb{E}[u_i \mid X_i] = 0.$$

3.2 Least Squares Objective

Estimate (β_0, β_1) by minimizing

$$S(b_0, b_1) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2.$$

First-order conditions:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \sum_{i=1}^{n} X_i \hat{u}_i = 0.$$

Proposition 3.1 (Closed-form OLS)**.**

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Proof. From $\sum \hat{u}_i = 0$ obtain $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$. Substitute into second normal equation and simplify. $\square$

3.3 Orthogonality Geometry

Key sample identities:

$$\sum_i \hat{u}_i = 0, \qquad \sum_i X_i \hat{u}_i = 0, \qquad \sum_i \hat{Y}_i \hat{u}_i = 0,$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. These generate decomposition of variation:

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i \hat{u}_i^2.$$

4 Finite-Sample Properties Under Exogeneity

4.1 Unbiasedness

Theorem 4.1. *If* $\mathbb{E}[u_i \mid X_i] = 0$ *and random sample assumptions hold,*

$$\mathbb{E}[\hat{\beta}_1 \mid X_1, \dots, X_n] = \beta_1, \qquad \mathbb{E}[\hat{\beta}_0 \mid X_1, \dots, X_n] = \beta_0.$$

Proof. Write

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}.$$

Conditioning on X, numerator has mean zero by linearity and $\mathbb{E}[u_i \mid X_i] = 0$. Intercept follows from $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$. □

4.2 Conditional Variance Under Homoskedasticity

If $\mathrm{Var}(u_i \mid X_i) = \sigma^2$ and errors are conditionally uncorrelated,

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2}.$$

Estimate with

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_i \hat{u}_i^2, \qquad SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_i (X_i - \bar{X})^2}}.$$

5 Asymptotic Theory for Simple OLS

5.1 Consistency

Theorem 5.1. *Under i.i.d. sampling,* $\mathbb{E}[u \mid X] = 0$*, and finite second moments,*

$$\hat{\beta}_1 \xrightarrow{p} \beta_1, \qquad \hat{\beta}_0 \xrightarrow{p} \beta_0.$$

Proof. Rewrite:

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_i (X_i - \mathbb{E}[X]) u_i - (\bar{X} - \mathbb{E}[X]) \bar{u}}{\frac{1}{n} \sum_i (X_i - \bar{X})^2}.$$

By LLN, numerator converges in probability to $\mathrm{Cov}(X, u) = 0$, denominator to $\mathrm{Var}(X) > 0$. □

5.2 Asymptotic Normality

Theorem 5.2. *Under i.i.d. sampling,* $\mathbb{E}[u \mid X] = 0$*, and finite moments,*

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0, \frac{\mathbb{E}[(X - \mathbb{E}[X])^2 u^2]}{\mathrm{Var}(X)^2}\right).$$

Sketch. Express

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_i (X_i - \mathbb{E}[X]) u_i + o_p(1)}{\frac{1}{n} \sum_i (X_i - \bar{X})^2}.$$

Apply CLT to numerator and LLN to denominator, then Slutsky. □

5.3 Homoskedastic Special Case

If $\mathbb{E}[u^2 \mid X] = \sigma^2$,

$$\mathbb{E}[(X - \mathbb{E}[X])^2 u^2] = \sigma^2 \operatorname{Var}(X),$$

so asymptotic variance simplifies to

$$\frac{\sigma^2}{\operatorname{Var}(X)}.$$

This highlights precision drivers:

1. larger n,
2. lower error variance,
3. more regressor variation.

6 Heteroskedasticity and Robust Standard Errors

6.1 Why Homoskedasticity is Optional for Consistency

OLS consistency relies on exogeneity, not constant variance. Therefore, heteroskedasticity does not by itself bias slope estimates.

6.2 Robust (Eicker-White) Variance in Simple Regression

In scalar regressor form,

$$\widehat{\operatorname{Var}}_{rob}(\hat{\beta}_1) = \frac{\sum_i (X_i - \bar{X})^2 \hat{u}_i^2}{\left(\sum_i (X_i - \bar{X})^2\right)^2}.$$

This is the empirical analog of asymptotic variance under arbitrary conditional variance patterns.

Remark 6.1. In applied workflow, robust standard errors are default unless strong design-based arguments justify homoskedastic formulas.

7 Goodness of Fit and Its Limits

Define

$$R^2 = 1 - \frac{\sum_i \hat{u}_i^2}{\sum_i (Y_i - \bar{Y})^2}.$$

In simple regression, R^2 equals squared sample correlation between X and Y.

Remark 7.1. High R^2 does not imply causal identification; low R^2 does not invalidate a causal effect estimate.

7.1 Adjusted R-Squared

In a regression with k non-intercept regressors and sample size n,

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{\mathrm{SSR}/(n-k-1)}{\mathrm{TSS}/(n-1)}.$$

Unlike R^2, adjusted R^2 can decrease when adding irrelevant controls.

7.2 Adding Regressors and Mechanical Fit Improvement

Proposition 7.2. *When regressors are added to an OLS model, R^2 weakly increases (never decreases).*

Proof. Let $\mathcal{B}_r$ be coefficient vectors in a restricted model and $\mathcal{B}_u$ in an unrestricted model with extra regressors. Then $\mathcal{B}_r \subseteq \mathcal{B}_u$, so

$$\min_{b \in \mathcal{B}_u} \sum_i (Y_i - x_i' b)^2 \le \min_{b \in \mathcal{B}_r} \sum_i (Y_i - x_i' b)^2.$$

Hence $\mathrm{SSR}_u \le \mathrm{SSR}_r$. Since $R^2 = 1 - \mathrm{SSR}/\mathrm{TSS}$ and TSS is fixed, $R_u^2 \ge R_r^2$. □

7.3 Partial R-Squared for a Block of Variables

Suppose the unrestricted model adds q variables to a restricted model. Define

$$R^2_{\text{partial}} = \frac{\mathrm{SSR}_r - \mathrm{SSR}_u}{\mathrm{SSR}_r}.$$

This is the fraction of remaining variation (after restricted controls) explained by the added block. If k_u is the number of non-intercept regressors in the unrestricted model,

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_u)/q}{\mathrm{SSR}_u/(n - k_u - 1)}, \qquad R^2_{\text{partial}} = \frac{qF}{qF + (n - k_u - 1)}.$$

This link helps interpret joint-significance tests in effect-size terms.

8 Inference for a Single Coefficient

8.1 t Statistic

For hypothesis $H_0 : \beta_1 = c$,

$$T = \frac{\hat{\beta}_1 - c}{SE(\hat{\beta}_1)}.$$

Large-sample two-sided 5% rule: reject if $|T| > 1.96$.

8.2 One-sided Alternatives

- $H_1 : \beta_1 > c$: reject if $T > z_{1-\alpha}$.
- $H_1 : \beta_1 < c$: reject if $T < -z_{1-\alpha}$.

8.3 Confidence Interval Duality

A $(1 - \alpha)$ CI is

$$\hat{\beta}_1 \pm z_{1-\alpha/2} SE(\hat{\beta}_1).$$

Reject $H_0 : \beta_1 = c$ in two-sided test iff c lies outside this interval.

8.4 Exact Finite-Sample t Inference Under Normal Errors

If, in addition, $u_i \mid X$ are i.i.d. normal and homoskedastic, then

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \Bigm| X \sim t_{n-2}.$$

So exact finite-sample t critical values can replace large-sample normal critical values. For large n, the difference is minor; for smaller samples, this correction matters.

9 Special Cases from 120B Additional Topics

9.1 Binary Regressor as Difference in Means

If $X \in \{0, 1\}$,

$$\hat{\beta}_1 = \bar{Y}_{X=1} - \bar{Y}_{X=0}, \qquad \hat{\beta}_0 = \bar{Y}_{X=0}.$$

Thus simple OLS recovers treated-control mean difference.

9.2 Units of Measurement

Rescaling X by $a \neq 0$ rescales slope by $1/a$; rescaling Y by b rescales slope by b. t statistics are invariant under linear rescaling if standard errors are transformed consistently.

9.3 Log Specifications

Common interpretations:

- level-log: $Y = \beta_0 + \beta_1 \ln X + u$; one-percent increase in X changes Y by about $0.01\beta_1$ units.
- log-level: $\ln Y = \beta_0 + \beta_1 X + u$; one-unit increase in X changes Y by approximately $100\beta_1\%$.
- log-log: $\ln Y = \beta_0 + \beta_1 \ln X + u$; β_1 is elasticity.

9.4 Classical Measurement Error in Regressor

Observed regressor $W = X + v$ with v independent noise. Regressing Y on W gives attenuation:

$$\operatorname{plim} \hat{\beta}_1^{(W)} = \beta_1 \frac{\operatorname{Var}(X)}{\operatorname{Var}(X) + \operatorname{Var}(v)}.$$

Bias is toward zero in absolute value when $\beta_1 \neq 0$.

Proof. With $Y = \beta_0 + \beta_1 X + u$,

$$\operatorname{plim} \hat{\beta}_1^{(W)} = \frac{\operatorname{Cov}(W, Y)}{\operatorname{Var}(W)} = \frac{\operatorname{Cov}(X + v, \beta_1 X + u)}{\operatorname{Var}(X) + \operatorname{Var}(v)} = \beta_1 \frac{\operatorname{Var}(X)}{\operatorname{Var}(X) + \operatorname{Var}(v)}.$$

□

9.5 Classical Measurement Error in Dependent Variable

Suppose we observe $Y^* = Y + \eta$ where η is independent of (X, u) and has mean zero. In regression of Y^* on X,

$$Y^* = \beta_0 + \beta_1 X + (u + \eta).$$

The new composite error keeps zero conditional mean given X, so slope consistency is preserved:

$$\text{plim } \hat{\beta}_1^{(Y^*)} = \beta_1.$$

But variance rises because $\text{Var}(u + \eta \mid X) = \text{Var}(u \mid X) + \text{Var}(\eta)$, so standard errors become larger.

10 Multiple Regression Foundations

10.1 Model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i.$$

Key identifying condition:

$$\mathbb{E}[u_i \mid X_{1i}, \ldots, X_{ki}] = 0.$$

10.2 Matrix Notation

Let $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times(k+1)}$ (first column ones), $\beta \in \mathbb{R}^{k+1}$, $u \in \mathbb{R}^n$:

$$y = X\beta + u.$$

OLS minimizes $(y - Xb)'(y - Xb)$.

Proposition 10.1 (Normal equations and estimator)**.** *If $X'X$ is invertible,*

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Proof. Differentiate objective:

$$\frac{\partial}{\partial b}(y - Xb)'(y - Xb) = -2X'(y - Xb).$$

Set to zero: $X'X\hat{b} = X'y$, solve by inversion. □

10.3 Interpretation of Partial Effects

β_j is the marginal effect of X_j on Y holding all other included regressors fixed. This partial-correlation logic is why multiple regression can reduce confounding from omitted variables.

11 Projection Algebra and Partialling Out

11.1 Hat Matrix and Residual Maker

With full-rank X, define

$$H = X(X'X)^{-1}X', \qquad M = I_n - H.$$

Then fitted values and residuals are

$$\hat{y} = Hy, \qquad \hat{u} = My.$$

Proposition 11.1. *H and M are symmetric and idempotent:*

$$H' = H,\ H^2 = H,\ M' = M,\ M^2 = M.$$

Also $X'\hat{u} = 0$, so residuals are orthogonal to every included regressor.

Proof. Symmetry and idempotence follow by direct multiplication using $(X'X)^{-1} = (X'X)^{-1\prime}$. Orthogonality:

$$X'\hat{u} = X'My = X'y - X'X(X'X)^{-1}X'y = 0.$$

□

11.2 Frisch-Waugh-Lovell (FWL) Theorem

Partition regressors as $X = [X_1\ X_2]$ where X_1 contains variables of interest and X_2 controls. Let

$$M_2 = I - X_2(X_2'X_2)^{-1}X_2'.$$

Then the OLS coefficient on X_1 from y on (X_1, X_2) is

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y.$$

So we can:

1. residualize y on X_2,
2. residualize X_1 on X_2,
3. run OLS of residualized y on residualized X_1.

Sketch. Normal equations imply

$$X_1'(y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2) = 0, \quad X_2'(y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2) = 0.$$

Solve second equation for $\hat{\beta}_2$, substitute into first, and factor out M_2. □

11.3 Interpretation Consequence

FWL clarifies what "holding controls fixed" means: coefficient identification comes only from variation in X_1 orthogonal to controls. If little residual variation remains, precision drops and standard errors rise.

12 Omitted Variable Bias

12.1 Two-Regressor OVB Formula

True model:

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u, \qquad \mathbb{E}[u \mid X, W] = 0.$$

If we regress Y only on X, then

$$\text{plim}\ \tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(X, W)}{\text{Var}(X)}.$$

Proof. By projection algebra,

$$\tilde{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(X, \beta_0 + \beta_1 X + \beta_2 W + u)}{\mathrm{Var}(X)}.$$

Since $\mathrm{Cov}(X,u) = 0$ under the long-model exogeneity,

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\mathrm{Cov}(X,W)}{\mathrm{Var}(X)}.$$

□

12.2 Sign Logic

Bias sign is determined by sign of $\beta_2 \,\mathrm{Cov}(X,W)$. Both "omitted variable affects outcome" and "omitted variable correlated with included regressor" are required for nonzero OVB.

13 Finite-Sample and Asymptotic Properties in Multiple Regression

13.1 Conditional Unbiasedness

Under $\mathbb{E}[u \mid X] = 0$ and full rank,

$$\mathbb{E}[\hat{\beta} \mid X] = \beta.$$

Because

$$\hat{\beta} = \beta + (X'X)^{-1}X'u,$$

then conditional expectation term vanishes.

13.2 Asymptotic Distribution

Let x_i denote row vector including intercept. Define

$$Q = \mathbb{E}[x_i x_i'], \qquad \Omega = \mathbb{E}[u_i^2 x_i x_i'].$$

Then

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1}).$$

Homoskedastic special case: $\Omega = \sigma^2 Q$, giving $\sigma^2 Q^{-1}$.

13.3 Robust Covariance Estimator

Sample analog:

$$\widehat{\mathrm{Var}}_{rob}(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'\right)(X'X)^{-1}.$$

Use this for robust t and Wald tests.

13.4 Homoskedastic Covariance in Matrix Form

If $\mathrm{Var}(u \mid X) = \sigma^2 I_n$, then

$$\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 (X'X)^{-1}.$$

This is exact finite-sample variance under conditional homoskedasticity.

13.5 Gauss-Markov Result (BLUE Statement)

Theorem 13.1. *Under linear model assumptions,* $\mathbb{E}[u \mid X] = 0$*, full rank, and* $\mathrm{Var}(u \mid X) = \sigma^2 I_n$*, the OLS estimator has the smallest conditional variance among all linear unbiased estimators.*

Sketch. Any linear unbiased estimator can be written $\tilde{\beta} = Cy$ with $CX = I$. Decompose

$$C = (X'X)^{-1}X' + D, \qquad DX = 0.$$

Then

$$\mathrm{Var}(\tilde{\beta} \mid X) - \mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 DD' \succeq 0.$$

So OLS is conditionally efficient in the linear-unbiased class. □

13.6 Perfect and Imperfect Multicollinearity

Perfect multicollinearity means one regressor is an exact linear combination of others, so $X'X$ is singular and OLS is not uniquely defined. Imperfect multicollinearity keeps identification but can inflate variances.

Proposition 13.2 (Variance inflation identity)**.** *Under homoskedasticity, for regressor* X_j *in a model with intercept and other controls,*

$$\mathrm{Var}(\hat{\beta}_j \mid X) = \frac{\sigma^2}{SST_j(1 - R_j^2)},$$

where $SST_j = \sum_i (X_{ji} - \bar{X}_j)^2$ *and* R_j^2 *is from regressing* X_j *on other regressors.*

Sketch. Apply FWL: coefficient on X_j equals slope from regression of $M_{-j}y$ on $M_{-j}X_j$. Homoskedastic scalar-OLS variance with residualized regressor yields denominator $\sum_i (M_{-j}X_j)_i^2 = \mathrm{SST}_j(1 - R_j^2)$. □

Remark 13.3. As $R_j^2 \to 1$, identifying variation in X_j disappears and $SE(\hat{\beta}_j)$ explodes.

14 Functional Form in Multiple Regression

14.1 Polynomials

Model with quadratic term:

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u.$$

Marginal effect of experience:

$$\frac{\partial \ln(wage)}{\partial exper} = \beta_2 + 2\beta_3 exper.$$

$\beta_3 < 0$ implies decreasing marginal effect.

14.2 Continuous-Continuous Interaction

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 (educ \cdot exper) + u.$$

Then

$$\frac{\partial \ln(wage)}{\partial exper} = \beta_2 + \beta_3 educ, \qquad \frac{\partial \ln(wage)}{\partial educ} = \beta_1 + \beta_3 exper.$$

Interaction sign indicates complementarity or substitutability in marginal effects.

14.3 Binary-Continuous Interaction

With female dummy D:

$$\ln(wage) = \beta_0 + \beta_1 D + \beta_2 educ + \beta_3(D \cdot educ) + u.$$

Group equations:

$$D = 0: \ \ln(wage) = \beta_0 + \beta_2 educ + u,$$
$$D = 1: \ \ln(wage) = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) educ + u.$$

So β_1 is intercept shift; β_3 is slope differential.

14.4 Categorical Regressors

For G groups, include $G - 1$ dummies plus intercept to avoid perfect multicollinearity (dummy variable trap).

14.5 Exact vs Approximate Log Interpretation

For model $\ln Y = \beta_0 + \beta_1 X + u$, a one-unit increase in X implies

$$\frac{Y_{\text{new}} - Y_{\text{old}}}{Y_{\text{old}}} = e^{\beta_1} - 1.$$

For small $|\beta_1|$, $e^{\beta_1} - 1 \approx \beta_1$, so the percent interpretation $100\beta_1\%$ is a first-order approximation.

14.6 Turning Point in Quadratic Models

In

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u, \quad \beta_2 \neq 0,$$

the slope is $\beta_1 + 2\beta_2 X$. Setting this to zero gives turning point

$$X^* = -\frac{\beta_1}{2\beta_2}.$$

If $\beta_2 < 0$, the relation is concave with a maximum at X^*; if $\beta_2 > 0$, convex with a minimum.

14.7 Centering in Interaction Models

In models with interaction $X_1 X_2$, centering variables as $\tilde{X}_j = X_j - \bar{X}_j$ can improve interpretability: the coefficient on $\tilde{X}_1$ becomes the effect of X_1 at mean X_2, and vice versa. Centering does not change fitted values or the interaction slope estimate.

15 Testing Linear Restrictions

15.1 Single Linear Combination

For null $H_0 : c'\beta = r$,

$$T = \frac{c'\hat{\beta} - r}{\sqrt{c'\widehat{\text{Var}}(\hat{\beta})c}}.$$

This includes tests like $H_0 : \beta_1 + \beta_2 = 1$.

15.2 Variance of Linear Combination

For two coefficients,

$$\operatorname{Var}(c_1\hat{\beta}_1 + c_2\hat{\beta}_2) = c_1^2 \operatorname{Var}(\hat{\beta}_1) + c_2^2 \operatorname{Var}(\hat{\beta}_2) + 2c_1c_2 \operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_2).$$

Ignoring covariance can materially misstate standard errors.

16 Joint Testing: Bonferroni and F/Wald

16.1 Bonferroni Approach

Test J restrictions $H_0 : \theta_j = 0$ using max t-statistic:

$$T_B = \max_{1 \le j \le J} |t_j|.$$

Reject if any $|t_j| > k_{\alpha/J}$.

Proposition 16.1 (Bonferroni size control)**.** *If each component test is level α/J, then joint rejection probability under H_0 is at most α.*

Proof. By union bound,

$$\mathbb{P}\left(\bigcup_{j=1}^{J} \{|t_j| > k_{\alpha/J}\}\right) \le \sum_{j=1}^{J} \mathbb{P}(|t_j| > k_{\alpha/J}) \le J \cdot \frac{\alpha}{J} = \alpha.$$

□

16.2 F/Wald Approach

For q linear restrictions $R\beta = r$ (with $R \in \mathbb{R}^{q\times(k+1)}$):

$$W = (R\hat{\beta} - r)'\big(R\widehat{\operatorname{Var}}(\hat{\beta})R'\big)^{-1}(R\hat{\beta} - r).$$

Asymptotically, $W \xrightarrow{d} \chi^2_q$. Equivalent F-form is $F = W/q$ with large-sample interpretation.

16.3 Overall Significance

The standard regression output F-statistic tests

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$

(excluding intercept). This is a joint relevance test, not a causal test.

16.4 Bonferroni Critical Values in Practice

For two-sided overall level α and J restrictions, each component test uses level α/J. Hence critical value is

$$z_{1-\alpha/(2J)}.$$

Examples used repeatedly in 120B:

- $\alpha = 0.05,\ J = 2$: critical value $\approx z_{0.9875} = 2.24$.
- $\alpha = 0.01,\ J = 2$: critical value $\approx z_{0.9975} = 2.807$.
- $\alpha = 0.01,\ J = 3$: critical value $\approx z_{0.99833} = 2.94$.

16.5 Bonferroni Joint Confidence Intervals

For two coefficients (β_a, β_b) with overall confidence $1-\alpha$, build

$$\hat{\beta}_a \pm z_{1-\alpha/4} SE(\hat{\beta}_a), \qquad \hat{\beta}_b \pm z_{1-\alpha/4} SE(\hat{\beta}_b).$$

By Bonferroni, simultaneous coverage probability is at least $1-\alpha$.

17 Worked Derivations and Examples

17.1 Example A: Difference-in-Means via OLS

If treatment indicator $X \in \{0,1\}$ and regression $Y = \beta_0 + \beta_1 X + u$,

$$\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0.$$

With random assignment, this is unbiased for ATE.

17.2 Example B: OVB Sign

Suppose true effect of class size on test score is negative ($\beta_1 < 0$). If funding W raises scores ($\beta_2 > 0$) and better-funded districts have smaller classes ($\text{Cov}(X, W) < 0$ where X is student-teacher ratio), then

$$\beta_2 \frac{\text{Cov}(X, W)}{\text{Var}(X)} < 0,$$

so short-regression slope is more negative than β_1 (downward bias).

17.3 Example C: Linear Restriction Test

Given estimates $\hat{\beta}_1 = 0.418$, $\hat{\beta}_2 = 0.631$, and covariance matrix entries

$$\widehat{\text{Var}}(\hat{\beta}_1) = 0.000916, \quad \widehat{\text{Var}}(\hat{\beta}_2) = 0.000913, \quad \widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.000246,$$

test $H_0 : \beta_1 + \beta_2 = 1$.

SE of sum:

$$SE(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{0.000916 + 0.000913 + 2(-0.000246)} \approx 0.037.$$

Statistic:

$$T = \frac{0.418 + 0.631 - 1}{0.037} \approx 1.32.$$

Two-sided p-value is around 0.19, so fail to reject.

17.4 Example D: Polynomial Marginal Effect

Suppose

$$\widehat{\ln(wage)} = 0.13 + 0.09 educ + 0.041 exper - 0.0007 exper^2.$$

Marginal effect of one more year of experience at $exper = 10$:

$$100 \times (0.041 - 2 \cdot 0.0007 \cdot 10) = 2.7\%.$$

At $exper = 30$, effect becomes near zero or negative, reflecting concavity and lifecycle patterns.

17.5 Example E: Bonferroni with Two Restrictions

Testing $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ jointly at overall 5% with Bonferroni:

$$\alpha^* = 0.05/2 = 0.025.$$

Two-sided critical value becomes about $z_{1-0.025/2} = z_{0.9875} \approx 2.24$ for each component statistic.

17.6 Example F: Bonferroni at 1% with Two Restrictions

Suppose we jointly test two coefficients at overall 1% level. Then each two-sided component test uses

$$\alpha^* = 0.01/2 = 0.005,$$

with per-tail probability 0.0025, giving critical value

$$z_{0.9975} \approx 2.807.$$

Decision rule: reject joint null if at least one component t-statistic exceeds 2.807 in absolute value.

17.7 Example G: Dummy Variable Trap

Suppose a wage regression includes intercept and four region dummies: Northeast, Midwest, South, West. Since these dummies sum to one for every observation,

$$1 = D_{\mathrm{NE}} + D_{\mathrm{MW}} + D_{\mathrm{S}} + D_{\mathrm{W}},$$

there is perfect multicollinearity. One region must be omitted as baseline. If West is omitted, each included dummy measures mean log-wage difference relative to West, holding other regressors fixed.

17.8 Example H: Scaling Transformations

Original model:

$$Y = \beta_0 + \beta_1 X + u.$$

Define transformed variables $Y^* = Y/1000$ and $X^* = 10X$. Then

$$Y^* = \frac{\beta_0}{1000} + \frac{\beta_1}{10000} X^* + \frac{u}{1000}.$$

So slope rescales deterministically. Hypothesis tests about the same economic effect are invariant once coefficients and standard errors are transformed consistently.

17.9 Example I: Testing a Nonlinear-and-Interaction Block

Consider

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 (educ \cdot exper) + u.$$

To test "no nonlinear or interaction effect," set

$$H_0 : \beta_3 = \beta_4 = 0.$$

This is a two-restriction joint test. Use either:

- Bonferroni with critical value 2.807 at overall 1%, or
- an $F/Wald$ test with $q = 2$ restrictions.

Rejecting implies at least one higher-order term contributes conditional explanatory power.

17.10 Example J: Measurement Error Contrast

If regressor error is classical ($W = X + v$), slope attenuates toward zero:

$$\text{plim } \hat{\beta}_1^{(W)} = \beta_1 \frac{\text{Var}(X)}{\text{Var}(X) + \text{Var}(v)}.$$

If dependent-variable error is classical ($Y^* = Y + \eta$), slope remains consistent but less precise. Hence "noise in X" is a bias issue; "noise in Y" is mainly a precision issue.

18 Extended Derivation Workbook

18.1 Finite-Sample OVB Decomposition

Suppose the true model is

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u, \qquad \mathbb{E}[u \mid X, W] = 0.$$

Define the sample linear projection of W on X:

$$W_i = \pi_0 + \pi_1 X_i + r_i, \qquad \sum_i X_i r_i = 0.$$

Substitute into the true model:

$$Y_i = (\beta_0 + \beta_2 \pi_0) + (\beta_1 + \beta_2 \pi_1) X_i + \beta_2 r_i + u_i.$$

Running short regression of Y on X alone gives

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \pi_1 + \frac{\sum_i (X_i - \bar{X})(\beta_2 r_i + u_i)}{\sum_i (X_i - \bar{X})^2}.$$

Since sample covariance of X and r is zero by construction, the r_i term drops:

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \pi_1 + \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}.$$

Taking probability limits yields the standard OVB expression using

$$\text{plim } \pi_1 = \frac{\text{Cov}(X, W)}{\text{Var}(X)}.$$

18.2 One Restriction: Wald, F, and t Equivalence

For a single restriction $H_0 : c'\beta = r$,

$$t = \frac{c'\hat{\beta} - r}{\sqrt{c'\widehat{\text{Var}}(\hat{\beta})c}}, \qquad W = \frac{(c'\hat{\beta} - r)^2}{c'\widehat{\text{Var}}(\hat{\beta})c}.$$

Hence

$$W = t^2.$$

If an F-statistic is reported as $F = W/q$, then for $q = 1$, $F = t^2$ exactly. So single-parameter tests are numerically identical across these representations.

18.3 Linear Combination Confidence Interval

To build a CI for $\theta = a'\beta$:

1. compute point estimate $\hat{\theta} = a'\hat{\beta}$,
2. compute standard error $SE(\hat{\theta}) = \sqrt{a'\widehat{\mathrm{Var}}(\hat{\beta})a}$,
3. report $\hat{\theta} \pm z_{1-\alpha/2}SE(\hat{\theta})$ (or t critical value in small samples).

This covers quantities like returns-to-schooling sums, policy contrasts, and average differential effects.

18.4 Delta Method for Log-Level Interpretation

In $\ln Y = \beta_0 + \beta_1 X + u$, exact percent effect of one-unit increase in X is

$$g(\beta_1) = 100(e^{\beta_1} - 1).$$

If $\hat{\beta}_1$ is asymptotically normal with variance V_{β_1}, then

$$\widehat{\mathrm{Var}}(g(\hat{\beta}_1)) \approx \left(100e^{\hat{\beta}_1}\right)^2 V_{\beta_1}.$$

So

$$SE\Big(100(e^{\hat{\beta}_1} - 1)\Big) \approx 100e^{\hat{\beta}_1}SE(\hat{\beta}_1).$$

This is useful when coefficients are not "small," where linear approximation $100\beta_1$ becomes rough.

18.5 Joint Hypotheses in Polynomial-Interaction Models

Consider

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4(educ \cdot exper) + u.$$

Common hypotheses:

1. no nonlinearity in experience: $H_0 : \beta_3 = 0$,
2. no interaction: $H_0 : \beta_4 = 0$,
3. no higher-order effects jointly: $H_0 : \beta_3 = \beta_4 = 0$.

Using separate 5% tests for (1) and (2) does not control joint size for (3). Bonferroni or F/Wald should be used for the joint claim.

18.6 Interpreting Non-Rejection Correctly

Failing to reject H_0 is not proof that H_0 is true. Typical reasons for non-rejection:

- true effect is near zero,
- sample is too small for adequate power,
- regressors are highly collinear,
- measurement error inflates variance.

Hence reporting effect sizes and confidence intervals is essential, not only p-values.

19 Assumption and Interpretation Checklist

19.1 Checklist

1. Is treatment/regressor variation plausibly exogenous?
2. Are key confounders controlled, or is design-based randomization available?
3. Are robust standard errors used when heteroskedasticity is plausible?
4. For nonlinear terms/interactions, are marginal effects evaluated at meaningful values?
5. For joint claims, is a joint test used (Bonferroni or F/Wald) instead of separate t tests?
6. Are causal statements separated from predictive fit metrics (R^2)?

19.2 Common Failure Modes

- interpreting correlation or fit as causation,
- omitting confounders and reading short-regression coefficients causally,
- testing multiple restrictions with unadjusted individual critical values,
- ignoring covariance terms in linear-combination standard errors,
- overinterpreting significance without effect-size context.

20 Proof Appendix

20.1 Frisch-Waugh-Lovell (Two Regressors, Scalar Form)

Model:

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u.$$

Regress X on $(1, W)$ and keep residual $\tilde{X}$. Regress Y on $(1, W)$ and keep residual $\tilde{Y}$. Then

$$\hat{\beta}_1 = \frac{\sum_i \tilde{X}_i \tilde{Y}_i}{\sum_i \tilde{X}_i^2}.$$

So multiple-regression partial effect equals simple-regression slope after partialling out controls.

20.2 Union-Bound Proof for Multiple Testing Control

For any events $A_1, \ldots, A_J$,

$$\mathbb{P}\left(\bigcup_{j=1}^{J} A_j\right) \le \sum_{j=1}^{J} \mathbb{P}(A_j).$$

Bonferroni is a direct corollary with A_j as individual false-rejection events.

20.3 Sandwich Variance Intuition

Asymptotically,

$$\sqrt{n}(\hat{\beta} - \beta) \approx Q_n^{-1} \cdot \frac{1}{\sqrt{n}} \sum_i x_i u_i.$$

Variance is thus

$$Q^{-1} \Omega Q^{-1}, \quad Q = \mathbb{E}[x_i x_i'], \quad \Omega = \mathbb{E}[u_i^2 x_i x_i'].$$

Replacing expectations by sample analogs gives robust covariance estimator.

20.4 R-Squared Equals Squared Correlation in Simple Regression

In simple OLS with intercept,

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_1 (X_i - \bar{X}).$$

Therefore

$$\text{ESS} = \sum_i (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_1^2 \sum_i (X_i - \bar{X})^2.$$

Using

$$\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2},$$

we obtain

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\left(\sum_i (X_i - \bar{X})(Y_i - \bar{Y})\right)^2}{\left(\sum_i (X_i - \bar{X})^2\right)\left(\sum_i (Y_i - \bar{Y})^2\right)} = \widehat{\text{Corr}}(X, Y)^2.$$

20.5 F Statistic from Restricted and Unrestricted SSR

For q linear restrictions and unrestricted model with k_u non-intercept regressors,

$$F = \frac{(\text{SSR}_r - \text{SSR}_u)/q}{\text{SSR}_u/(n - k_u - 1)}.$$

This compares average fit gain per restriction to unrestricted residual variance. Large values imply restrictions are incompatible with the data.

20.6 Consistency Under Classical Error in Y

If observed outcome is $Y^* = Y + \eta$ with $\mathbb{E}[\eta \mid X] = 0$ and $\text{Cov}(\eta, X) = 0$, then regression error becomes $u^* = u + \eta$ and

$$\mathbb{E}[u^* \mid X] = \mathbb{E}[u \mid X] + \mathbb{E}[\eta \mid X] = 0.$$

Hence OLS exogeneity still holds and slope consistency follows by the same LLN argument as the baseline model.

21 Practice Set with Solution Sketches

21.1 Q1: Causal Identification

Why does random assignment imply $\mathbb{E}[u \mid X] = 0$ in the binary treatment regression representation?

Sketch: random assignment makes X independent of $u(1)$ and $u(0)$. With

$$u = u(0) + X\big(u(1) - u(0)\big),$$

this implies $\mathbb{E}[u \mid X] = 0$.

21.2 Q2: OLS Derivation

Derive normal equations from SSR minimization and solve for $(\hat{\beta}_0, \hat{\beta}_1)$.

Sketch: differentiate SSR with respect to both parameters, set to zero, solve linear system.

21.3 Q3: OVB Mechanics

Given true model $Y = \beta_0 + \beta_1 X + \beta_2 W + u$, prove short-regression bias formula and discuss sign in a class-size application.

21.4 Q4: Robust vs Conventional SE

Explain why robust and homoskedastic SE may differ even when point estimates are identical.

Sketch: both use same $\hat{\beta}$, but variance estimators differ in weighting residual squares.

21.5 Q5: Joint Restriction

How would you test $H_0 : \beta_{educ^2} = \beta_{exper^2} = 0$ in a wage regression with quadratic terms?

Sketch: use $R\beta = r$ with two restrictions and compute F/Wald statistic.

21.6 Q6: Interaction Interpretation

In model $\ln(wage) = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 (female \cdot educ) + u$, interpret β_3.

Sketch: difference in returns to education between women and men.

21.7 Q7: Multiple-Testing Logic

Why can testing each restriction at 5% inflate overall false-rejection probability in joint hypothesis testing?

Sketch: union probability across multiple tests exceeds each marginal test size; use Bonferroni or F/Wald.

21.8 Q8: Elasticity Interpretation

In log-log model, what does slope 0.8 mean?

Sketch: a 1% increase in X is associated with 0.8% increase in Y.

21.9 Q9: Measurement Error

Why does classical error in X generally bias slope toward zero?

Sketch: noise inflates denominator variance but not covariance signal proportionately.

21.10 Q10: Causal Language Discipline

A regression finds statistically significant coefficient with high R^2. What extra condition is needed for causal interpretation?

Sketch: exogeneity/identification condition such as random assignment or valid control strategy ensuring $\mathbb{E}[u \mid X, controls] = 0$.

21.11 Q11: Adjusted R-Squared

Why can adjusted R^2 decrease after adding a regressor even though R^2 cannot?

Sketch: R^2 is mechanical in-sample fit, while adjusted R^2 penalizes degrees of freedom through $(n - k - 1)$.

21.12 Q12: FWL Computation

Describe a three-step algorithm to estimate the coefficient on X_1 in Y on (X_1, X_2) without running the full regression directly.

Sketch: residualize Y on X_2, residualize X_1 on X_2, regress residualized Y on residualized X_1.

21.13 Q13: Multicollinearity

Why can two individually insignificant coefficients become jointly significant in an F test?

Sketch: high covariance among estimates can inflate individual SEs while the joint linear combination is still precisely estimated.

21.14 Q14: Dummy Baseline

In a model with 5 mutually exclusive categories and an intercept, how many dummies can be included, and how are coefficients interpreted?

Sketch: include 4 dummies; each measures difference relative to omitted baseline category.

21.15 Q15: Bonferroni 1% Rule

For three restrictions at overall 1%, what two-sided critical value should each t test use?

Sketch: use $z_{1-0.01/(2\cdot3)} \approx z_{0.99833} \approx 2.94$.

21.16 Q16: Measurement Error

Give one empirical symptom that can suggest attenuation bias from classical error in X.

Sketch: estimates move away from zero when replacing noisy regressor with a more reliable measure or instrument.

22 Derivation and Proof Summary

1. Potential outcomes produce a regression form where slope can equal ATE under random assignment.
2. OLS derives from orthogonality and projection, not ad hoc fitting.
3. Exogeneity controls bias; homoskedasticity controls only variance simplification.

4. OVB has transparent sign and magnitude formula.
5. Linear restrictions require full covariance-aware testing.
6. Joint claims need joint tests (Bonferroni/F/Wald), not isolated t statistics.

23 Bridge to 120C

These tools feed directly into advanced topics: panel methods, causal designs beyond selection-on-observables, instrumental variables, and quasi-experimental identification.