

ECON 120C: Synthesized Lecture Notes

Binary Outcomes, IV, Panel Data, and DiD

Compiled from ECON 120C lecture sequence

Contents

1	Scope and Sequence	5
2	Potential Outcomes and Algebra	5
2.1	Core Setup	5
2.2	Average Causal Parameters	5
2.3	Selection-Bias Decomposition	5
2.4	Random Assignment as Identification	5
2.5	Regression Representation	6
2.6	Potential-Outcomes Form of Difference-in-Differences	6
3	Binary Outcome Models	6
3.1	Setup	6
3.2	Linear Probability Model	6
3.3	Implications for Inference	7
3.4	Interpretation and Practical Limits	7
3.5	Probit and Logit Models	7
3.6	Marginal Effects in Nonlinear Binary Models	7
3.7	Logit Odds Interpretation	7
3.8	Goodness of Fit for Binary Models	8
4	Endogeneity and Why OLS Fails	8
4.1	Sources of Endogeneity	8
4.2	Probability Limit of OLS with One Regressor	8
5	Instrumental Variables with One Endogenous Regressor	8
5.1	Model and Instrument Conditions	8
5.2	Just-Identified IV Estimator	9
5.3	Wald Estimator for Binary Instrument	9
5.4	Reduced Form and First Stage	9
6	Two-Stage Least Squares in Multiple Regression	9
6.1	General Setup	9
6.2	Two-Stage Algorithm	10
6.3	Matrix Formula	10
6.4	Asymptotic Distribution and Robust Variance	10
6.5	Exactly Identified and Overidentified Cases	10
6.6	Overidentification Logic	10

7	Local Causal Interpretation of IV	10
7.1	Potential Outcomes with Instrumental Assignment	10
7.2	Compliance Types	11
7.3	Key Assumptions for Wald Interpretation	11
7.4	Interpretation Discipline	11
8	Weak Instruments	11
8.1	Why Weak Relevance is Dangerous	11
8.2	First-Stage Diagnostics	11
8.3	Partial R-Squared	12
8.4	Inference Under Weak Identification	12
8.5	Testing Endogeneity with OLS and IV	12
9	Panel Data and Fixed Effects	12
9.1	One-Way Fixed Effects Model	12
9.2	Two-Way Fixed Effects Model	12
9.3	Within Transformation	12
9.4	Dummy Variable (LSDV) Equivalence	13
9.5	What FE Cannot Identify	13
9.6	Counting Coefficients in Two-Way FE	13
9.7	First-Difference Estimator	13
9.8	FE and FD with Two Periods	13
9.9	Random Effects and FE Comparison	13
9.10	Unbalanced Panels	13
9.11	Lagged Outcomes and Dynamic Bias	13
10	Clustered Inference for Panel Data	14
10.1	Why Usual Robust SE Can Still Fail	14
10.2	Cluster-Robust Variance	14
10.3	Interpretation	14
10.4	Practical Rule	14
11	Difference-in-Differences	14
11.1	Canonical 2x2 Setup	14
11.2	Regression Implementation	14
11.3	Parallel Trends Assumption	15
11.4	Repeated Cross Sections vs Panels	15
11.5	DiD with Covariates	15
11.6	Two-Way FE DiD Form	15
11.7	Event-Study Representation	15
11.8	Pre-Trend Checks	15
11.9	Threats to DiD Validity	15
11.10	Difference-in-Difference-in-Differences	16
12	Worked Derivations and Empirical Examples	16
12.1	Example A: Mortgage Denial Probit	16
12.2	Example B: O-Ring Failure and Binary Models	16
12.3	Example C: Instrument Validity for Returns to Education	16

12.4	Example D: Two-Group Restaurant DiD	17
12.5	Example E: Cigarette Tax Natural Experiment	17
12.6	Example F: Panel FE and Policy Effects	17
12.7	Example G: Counting FE Parameters	17
12.8	Example H: Region FE Restriction Test	17
12.9	Example I: Weak-IV Sensitivity	17
12.10	Example J: Difference Between FE and Pooled OLS	18
13	Extended Derivation Workbook	18
13.1	Deriving IV from Orthogonality Conditions	18
13.2	Generated Regressor View of 2SLS	18
13.3	Bias Direction Under Weak Instruments	18
13.4	Within Estimator as OLS on Residualized Data	18
13.5	Two-Way FE Residualization	18
13.6	DiD in First-Difference Form	19
13.7	Standard Error Choice in DiD	19
13.8	Linear Restrictions in FE Panels	19
13.9	How Unclustered SE Can Overstate Precision	19
13.10	DiD Decomposition with Covariates	19
13.11	ATT Interpretation in DiD	19
14	Assumption and Interpretation Checklist	19
14.1	Binary Outcome Models	19
14.2	IV Design	20
14.3	Panel and DiD	20
15	Common Failure Modes	20
16	Proof Appendix	20
16.1	LPM Conditional Variance Identity	20
16.2	IV Orthogonality Proof	20
16.3	FWL Proof Sketch for FE	21
16.4	DiD Coefficient Identity	21
16.5	Cluster-Robust Intuition	21
16.6	FE and FD Equivalence When T Equals Two	21
16.7	LATE Decomposition Identity	21
16.8	Serial Correlation and DiD Variance	21
17	Practice Set with Solution Sketches	22
17.1	Q1: LPM Heteroskedasticity	22
17.2	Q2: Nonlinear Marginal Effects	22
17.3	Q3: OLS Endogeneity Bias	22
17.4	Q4: IV Consistency	22
17.5	Q5: First Stage Strength	22
17.6	Q6: Overidentification	22
17.7	Q7: FE Identification	22
17.8	Q8: Cluster SE	22
17.9	Q9: DiD Estimand	22

17.10Q10: Parallel Trends	22
17.11Q11: Repeated Cross Section DiD	23
17.12Q12: Binary Model Choice	23
17.13Q13: Policy Interpretation	23
17.14Q14: DiD Standard Errors	23
17.15Q15: Instrument Screening	23
17.16Q16: FE vs Pooled OLS	23
17.17Q17: LATE	23
17.18Q18: Monotonicity	23
17.19Q19: FE vs RE Choice	23
17.20Q20: Dynamic Panel	23
17.21Q21: Event Study	24
17.22Q22: DiD With Controls	24
17.23Q23: Overidentification	24
17.24Q24: Cluster Level	24
18 Derivation and Proof Summary	24
19 Bridge to Integrated ECON 120 Notes	24

1 Scope and Sequence

These notes synthesize the ECON 120C arc into one coherent document:

1. binary-outcome models (LPM, probit, logit),
2. endogeneity and instrumental variables,
3. panel-data fixed effects and clustered inference,
4. difference-in-differences design and regression implementation.

The focus is derivation and proof: each estimator is tied to assumptions, then connected to implementation and interpretation.

2 Potential Outcomes and Algebra

2.1 Core Setup

Let treatment be $D_i \in \{0, 1\}$ with potential outcomes $Y_i(1)$ and $Y_i(0)$. Observed outcome satisfies

$$Y_i = Y_i(0) + D_i(Y_i(1) - Y_i(0)).$$

Define unit-level treatment effect

$$\tau_i = Y_i(1) - Y_i(0).$$

2.2 Average Causal Parameters

$$ATE = \mathbb{E}[\tau_i], \quad ATT = \mathbb{E}[\tau_i \mid D_i = 1], \quad ATU = \mathbb{E}[\tau_i \mid D_i = 0].$$

Only one potential outcome is observed per unit, so these objects are not directly observable without additional assumptions.

2.3 Selection-Bias Decomposition

Difference in observed means can be decomposed as

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = ATT + \left(\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0] \right).$$

The second term is selection bias in untreated potential outcomes.

Equivalent decomposition around ATE:

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = ATE + \underbrace{\left(\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0] \right)}_{\text{selection on baseline}} + \underbrace{\left(\mathbb{E}[\tau_i \mid D_i = 1] - ATE \right)}_{\text{selection on gains}}.$$

2.4 Random Assignment as Identification

If

$$D_i \perp (Y_i(1), Y_i(0)),$$

then

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] = ATE.$$

So random assignment removes both baseline-selection and gain-selection components.

2.5 Regression Representation

Write

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

with

$$\begin{aligned}\beta_0 &= \mathbb{E}[Y_i(0)], & \beta_1 &= \mathbb{E}[\tau_i] = ATE, \\ u_i &= (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i(\tau_i - \mathbb{E}[\tau_i]).\end{aligned}$$

Then

$$\mathbb{E}[u_i | D_i] = 0$$

is exactly the condition ruling out selection terms above.

2.6 Potential-Outcomes Form of Difference-in-Differences

With two periods $t \in \{0, 1\}$ and group indicator $G_i \in \{0, 1\}$,

$$Y_{it} = Y_{it}(0) + G_i \cdot Post_t \cdot (Y_{it}(1) - Y_{it}(0)).$$

DiD identifies

$$ATT_{post} = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) | G_i = 1]$$

under parallel trends in untreated outcomes:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) | G_i = 0].$$

This is the potential-outcomes algebra behind the regression interaction term in DiD.

3 Binary Outcome Models

3.1 Setup

Let $Y_i \in \{0, 1\}$ be a binary outcome and $X_i \in \mathbb{R}^k$ regressors. Define conditional success probability

$$p(x) = \mathbb{P}(Y_i = 1 | X_i = x) = \mathbb{E}[Y_i | X_i = x].$$

Because Y_i is binary, modeling $\mathbb{E}[Y_i | X_i]$ is equivalent to modeling a probability.

3.2 Linear Probability Model

The linear probability model (LPM) specifies

$$Y_i = \beta_0 + X_i' \beta + u_i, \quad \mathbb{E}[u_i | X_i] = 0.$$

Then

$$\mathbb{P}(Y_i = 1 | X_i) = \mathbb{E}[Y_i | X_i] = \beta_0 + X_i' \beta.$$

Proposition 3.1 (LPM error variance). *In the LPM,*

$$\text{Var}(u_i | X_i) = p(X_i)(1 - p(X_i)),$$

so conditional variance depends on X_i and is generally heteroskedastic.

Proof. Since $u_i = Y_i - p(X_i)$,

$$\text{Var}(u_i | X_i) = \text{Var}(Y_i | X_i) = \mathbb{E}[Y_i^2 | X_i] - \mathbb{E}[Y_i | X_i]^2.$$

For binary Y_i , $Y_i^2 = Y_i$, so

$$\text{Var}(u_i | X_i) = p(X_i) - p(X_i)^2 = p(X_i)(1 - p(X_i)).$$

□

3.3 Implications for Inference

OLS estimates of the LPM slope can still be consistent under exogeneity, but homoskedastic standard errors are invalid in general. Robust standard errors are therefore the default.

3.4 Interpretation and Practical Limits

In the LPM, β_j is the change in predicted probability from a one-unit increase in X_j , holding other regressors fixed. Potential drawback: predicted values \hat{p}_i can lie outside $[0, 1]$, especially in tails.

3.5 Probit and Logit Models

To keep probabilities in $[0, 1]$, use nonlinear links:

$$\mathbb{P}(Y_i = 1 | X_i) = \Phi(X_i' \beta) \quad (\text{probit}),$$

$$\mathbb{P}(Y_i = 1 | X_i) = \Lambda(X_i' \beta) = \frac{1}{1 + e^{-X_i' \beta}} \quad (\text{logit}).$$

Both are estimated by maximum likelihood.

3.6 Marginal Effects in Nonlinear Binary Models

For continuous regressor X_{ij} ,

$$\frac{\partial \mathbb{P}(Y_i = 1 | X_i)}{\partial X_{ij}} = \phi(X_i' \beta) \beta_j \quad (\text{probit}),$$

$$\frac{\partial \mathbb{P}(Y_i = 1 | X_i)}{\partial X_{ij}} = \Lambda(X_i' \beta) (1 - \Lambda(X_i' \beta)) \beta_j \quad (\text{logit}).$$

Hence marginal effects vary with covariate values.

Common summaries:

- Marginal effect at means (MEM),
- Average marginal effect (AME): average of individual marginal effects.

3.7 Logit Odds Interpretation

In a logit model,

$$\log \left(\frac{p_i}{1 - p_i} \right) = X_i' \beta,$$

so a one-unit increase in X_{ij} changes log-odds by β_j , holding other regressors fixed.

3.8 Goodness of Fit for Binary Models

R^2 from OLS is often less informative for binary outcomes. Common alternatives:

- pseudo- R^2 (e.g., McFadden),
- classification tables,
- ROC/AUC summaries.

4 Endogeneity and Why OLS Fails

4.1 Sources of Endogeneity

The regressor can correlate with the error due to:

1. omitted variables,
2. simultaneity/reverse causality,
3. measurement error in regressors.

4.2 Probability Limit of OLS with One Regressor

Suppose

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

with $\text{Cov}(X_i, u_i) \neq 0$. Then

$$\text{plim } \hat{\beta}_1^{OLS} = \beta_1 + \frac{\text{Cov}(X_i, u_i)}{\text{Var}(X_i)}.$$

This is the same algebra as omitted-variable bias: if X carries endogenous variation, OLS cannot isolate causal variation.

5 Instrumental Variables with One Endogenous Regressor

5.1 Model and Instrument Conditions

Let

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

where X_i is endogenous. A candidate instrument Z_i must satisfy:

1. **Relevance:** $\text{Cov}(Z_i, X_i) \neq 0$,
2. **Exogeneity:** $\text{Cov}(Z_i, u_i) = 0$.

5.2 Just-Identified IV Estimator

The sample IV slope is

$$\hat{\beta}_1^{IV} = \frac{\sum_i (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_i (Z_i - \bar{Z})(X_i - \bar{X})}.$$

Population analog:

$$\beta_1 = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)}.$$

Theorem 5.1 (Consistency of just-identified IV). *If relevance and exogeneity hold and moments are finite,*

$$\hat{\beta}_1^{IV} \xrightarrow{p} \beta_1.$$

Proof. By LLN,

$$\text{plim } \hat{\beta}_1^{IV} = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)}.$$

Using $Y_i = \beta_0 + \beta_1 X_i + u_i$,

$$\text{Cov}(Z_i, Y_i) = \beta_1 \text{Cov}(Z_i, X_i) + \text{Cov}(Z_i, u_i).$$

Exogeneity gives $\text{Cov}(Z_i, u_i) = 0$, so ratio equals β_1 . □

5.3 Wald Estimator for Binary Instrument

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

$$\beta_1 = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[X_i | Z_i = 1] - \mathbb{E}[X_i | Z_i = 0]}.$$

This is the Wald ratio form.

5.4 Reduced Form and First Stage

Write

$$X_i = \pi_0 + \pi_1 Z_i + v_i \quad (\text{first stage}),$$

$$Y_i = \delta_0 + \delta_1 Z_i + e_i \quad (\text{reduced form}).$$

Then under just-identification,

$$\beta_1 = \frac{\delta_1}{\pi_1}.$$

6 Two-Stage Least Squares in Multiple Regression

6.1 General Setup

Partition regressors into endogenous X_1 and exogenous X_2 :

$$y = X_1 \beta_1 + X_2 \beta_2 + u.$$

Let instruments be $Z = [X_2 \ W]$, where W are excluded instruments.

6.2 Two-Stage Algorithm

1. Regress each column of X_1 on Z to obtain fitted values \hat{X}_1 .
2. Regress y on (\hat{X}_1, X_2) by OLS.

6.3 Matrix Formula

Define projection onto instrument space:

$$P_Z = Z(Z'Z)^{-1}Z'$$

Then 2SLS estimator is

$$\hat{\beta}^{2SLS} = (X'P_ZX)^{-1}X'P_Zy,$$

where $X = [X_1 \ X_2]$.

6.4 Asymptotic Distribution and Robust Variance

Under standard conditions,

$$\sqrt{n}(\hat{\beta}^{2SLS} - \beta) \xrightarrow{d} N(0, V).$$

A heteroskedasticity-robust estimator is

$$\widehat{\text{Var}}(\hat{\beta}^{2SLS}) = (X'P_ZX)^{-1} (X'P_Z\hat{\Omega}P_ZX) (X'P_ZX)^{-1},$$

where $\hat{\Omega}$ is typically diagonal with entries \hat{u}_i^2 .

6.5 Exactly Identified and Overidentified Cases

If number of excluded instruments equals number of endogenous regressors, model is exactly identified. If instruments exceed endogenous regressors, model is overidentified and allows additional validity checks.

6.6 Overidentification Logic

When there are extra instruments, all valid instruments should imply the same structural coefficient. Large disagreements in overidentification tests suggest at least one instrument may violate exogeneity.

7 Local Causal Interpretation of IV

7.1 Potential Outcomes with Instrumental Assignment

Let $D_i \in \{0, 1\}$ be treatment and $Z_i \in \{0, 1\}$ instrument. Potential treatment status: $D_i(1), D_i(0)$. Potential outcomes under treatment: $Y_i(1), Y_i(0)$. Observed outcomes:

$$D_i = D_i(Z_i), \quad Y_i = Y_i(0) + D_i(Y_i(1) - Y_i(0)).$$

7.2 Compliance Types

Units are classified by $(D_i(1), D_i(0))$:

- compliers: $(1, 0)$,
- always-takers: $(1, 1)$,
- never-takers: $(0, 0)$,
- defiers: $(0, 1)$.

7.3 Key Assumptions for Wald Interpretation

For binary instrument:

1. instrument exogeneity: $Z \perp (Y(1), Y(0), D(1), D(0))$,
2. exclusion: Z affects Y only through D ,
3. relevance: $\mathbb{P}(D = 1 \mid Z = 1) \neq \mathbb{P}(D = 1 \mid Z = 0)$,
4. monotonicity: $D(1) \geq D(0)$ (no defiers).

Theorem 7.1 (Wald equals LATE). *Under assumptions above,*

$$\frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]} = \mathbb{E}[Y(1) - Y(0) \mid \text{complier}].$$

Sketch. Write reduced-form numerator as average treatment-effect contribution across compliance types. Always-takers and never-takers cancel in the difference across Z . Monotonicity removes defier terms, leaving complier average gain multiplied by complier share. Divide by first-stage denominator, which equals complier share. \square

7.4 Interpretation Discipline

IV with heterogeneous effects generally identifies a local effect for compliers, not the population ATE. Policy extrapolation therefore requires assumptions about external validity across compliance groups.

8 Weak Instruments

8.1 Why Weak Relevance is Dangerous

If $\text{Corr}(Z, X)$ is small, IV estimates become unstable and can be biased toward OLS in finite samples.

8.2 First-Stage Diagnostics

For a single endogenous regressor, first-stage F-statistic on excluded instruments is a common diagnostic. A widely used rule of thumb flags weakness when $F < 10$.

8.3 Partial R-Squared

First-stage partial R^2 of excluded instruments measures incremental explanatory power for endogenous regressors after controlling for included exogenous variables. Small partial R^2 means weak identifying variation.

8.4 Inference Under Weak Identification

Conventional Wald intervals may have severe size distortion under weak instruments. Robust alternatives include Anderson-Rubin style tests and weak-IV robust confidence sets.

8.5 Testing Endogeneity with OLS and IV

When OLS and IV estimates differ materially, one interpretation is endogeneity in OLS. A Hausman-type statistic compares coefficient differences to the variance of those differences. Rejection supports using IV for consistency; non-rejection does not prove exogeneity, especially under weak instruments.

9 Panel Data and Fixed Effects

9.1 One-Way Fixed Effects Model

For entities $i = 1, \dots, N$ and periods $t = 1, \dots, T$,

$$Y_{it} = \alpha_i + X'_{it}\beta + u_{it}.$$

α_i captures time-invariant unobserved heterogeneity.

9.2 Two-Way Fixed Effects Model

$$Y_{it} = \alpha_i + \lambda_t + X'_{it}\beta + u_{it},$$

where λ_t are common time shocks.

9.3 Within Transformation

For one-way FE, entity-demean:

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)' \beta + (u_{it} - \bar{u}_i).$$

α_i drops out, allowing consistent estimation of β if

$$\mathbb{E}[u_{it} \mid X_{i1}, \dots, X_{iT}, \alpha_i] = 0.$$

Theorem 9.1 (FE consistency under strict exogeneity). *Under strict exogeneity and rank conditions,*

$$\hat{\beta}^{FE} \xrightarrow{p} \beta$$

as $N \rightarrow \infty$ with fixed T .

9.4 Dummy Variable (LSDV) Equivalence

Including entity dummies and running OLS yields same slope estimates as within transformation. This follows from Frisch-Waugh-Lovell: partialling out entity means in both y and X is algebraically equivalent to demeaning.

9.5 What FE Cannot Identify

Any regressor constant within entity over time is perfectly collinear with entity fixed effects and is not identified in one-way FE. With two-way FE, regressors that vary only by time are also absorbed by λ_t .

9.6 Counting Coefficients in Two-Way FE

With intercept, k slope regressors, N entities, and T periods,

$$\text{parameters} = 1 + k + (N - 1) + (T - 1) = k + N + T - 1.$$

This count appears in panel model specification exercises and helps detect dummy-variable traps.

9.7 First-Difference Estimator

Differencing adjacent periods removes entity fixed effects:

$$\Delta Y_{it} = Y_{it} - Y_{i,t-1} = \beta'(X_{it} - X_{i,t-1}) + (u_{it} - u_{i,t-1}).$$

With strict exogeneity, OLS on differenced data is consistent.

9.8 FE and FD with Two Periods

When $T = 2$, FE and FD use identical transformed variation up to scale, so slope estimates coincide. With $T > 2$, FE and FD can differ in efficiency depending on serial-correlation structure.

9.9 Random Effects and FE Comparison

Random effects (RE) assumes entity effects are uncorrelated with regressors:

$$\text{Cov}(\alpha_i, X_{it}) = 0.$$

If this fails, RE is inconsistent while FE remains consistent under strict exogeneity. In practice, FE is often the safer baseline when omitted time-invariant heterogeneity is plausible.

9.10 Unbalanced Panels

Most empirical panels are unbalanced (different observation windows across entities). FE and cluster-robust inference still apply, but interpretation should note changing sample composition over time.

9.11 Lagged Outcomes and Dynamic Bias

Model

$$Y_{it} = \rho Y_{i,t-1} + \alpha_i + X'_{it}\beta + u_{it}$$

cannot be estimated consistently by standard FE when T is short because transformed lagged outcomes correlate with transformed error (Nickell bias).

10 Clustered Inference for Panel Data

10.1 Why Usual Robust SE Can Still Fail

Heteroskedasticity-robust SE allow arbitrary variance differences but often assume no within-cluster dependence. In panel settings, errors are commonly serially correlated within entity, invalidating conventional robust SE.

10.2 Cluster-Robust Variance

For clustering by entity g ,

$$\widehat{\text{Var}}_{cl}(\hat{\beta}) = (X'X)^{-1} \left(\sum_{g=1}^G X'_g \hat{u}_g \hat{u}'_g X_g \right) (X'X)^{-1},$$

with finite-sample corrections applied in software.

10.3 Interpretation

Clustering allows arbitrary correlation and heteroskedasticity inside each cluster, but assumes independence across clusters.

10.4 Practical Rule

Cluster at the level where treatment or policy varies and where serial dependence is plausible. In state-year policy panels, clustering by state is usually the baseline.

11 Difference-in-Differences

11.1 Canonical 2x2 Setup

Define treatment-group indicator $D_i \in \{0, 1\}$ and post indicator $Post_t \in \{0, 1\}$. Group-time means:

$$\mu_{d,p} = \mathbb{E}[Y \mid D = d, Post = p].$$

DiD estimand:

$$\Delta_{DiD} = (\mu_{1,1} - \mu_{1,0}) - (\mu_{0,1} - \mu_{0,0}).$$

11.2 Regression Implementation

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \cdot Post_t) + u_{it}.$$

Then $\beta_3 = \Delta_{DiD}$.

Proposition 11.1 (DiD-regression equivalence). *In the 2x2 design above, OLS coefficient on $D_i \cdot Post_t$ equals the difference-in-differences of cell means.*

Proof. Take fitted means from the regression:

$$\mathbb{E}[Y \mid D = 0, Post = 0] = \beta_0,$$

$$\mathbb{E}[Y \mid D = 1, Post = 0] = \beta_0 + \beta_1,$$

$$\mathbb{E}[Y \mid D = 0, Post = 1] = \beta_0 + \beta_2,$$

$$\mathbb{E}[Y \mid D = 1, Post = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3.$$

Substitute into DiD expression; all terms cancel except β_3 . □

11.3 Parallel Trends Assumption

The core identifying assumption is

$$\mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) \mid D_i = 1] = \mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) \mid D_i = 0].$$

Without this, β_3 mixes treatment effect with differential untreated trends.

11.4 Repeated Cross Sections vs Panels

DiD does not require following the same individuals over time. Repeated cross sections can identify DiD if group definitions are consistent and parallel-trends assumptions are credible at group level.

11.5 DiD with Covariates

A common extension is

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \cdot Post_t) + W_{it}' \gamma + u_{it},$$

where W_{it} improves precision and can absorb compositional changes.

11.6 Two-Way FE DiD Form

With many periods and groups:

$$Y_{it} = \alpha_i + \lambda_t + \beta Treat_{it} + u_{it}.$$

This is a generalization of DiD that relies on common-trend style assumptions conditional on fixed effects.

11.7 Event-Study Representation

For staggered or multi-period treatment, a common diagnostic is

$$Y_{it} = \alpha_i + \lambda_t + \sum_{\ell \neq -1} \theta_\ell \mathbf{1}\{EventTime_{it} = \ell\} + u_{it},$$

where leads ($\ell < 0$) check pre-treatment differential trends and lags ($\ell \geq 0$) trace dynamics.

11.8 Pre-Trend Checks

Failing to reject lead coefficients is not proof of parallel trends, but large pre-treatment deviations weaken DiD credibility. Graphing untreated period means for treatment and control groups remains a core diagnostic.

11.9 Threats to DiD Validity

Common threats:

- composition changes across periods,
- spillovers from treated to control units,
- anticipatory behavior before treatment date,
- differential shocks unrelated to treatment.

11.10 Difference-in-Difference-in-Differences

Adding a third difference can difference out one additional confounder dimension:

$$DDD = (\Delta_T - \Delta_C)_{group A} - (\Delta_T - \Delta_C)_{group B}.$$

This is useful when two-way differencing still leaves a structured omitted trend.

12 Worked Derivations and Empirical Examples

12.1 Example A: Mortgage Denial Probit

Estimated model:

$$\mathbb{P}(\text{deny} = 1 \mid \text{ratio}, \text{black}) = \Phi(-2.26 + 2.74 \text{ratio} + 0.71 \text{black}).$$

For a Black applicant with $\text{ratio} = 0.40$:

$$z = -2.26 + 2.74(0.40) + 0.71 = -0.454,$$

$$\hat{p} = \Phi(-0.454) \approx 0.325.$$

If ratio decreases to 0.30:

$$z = -0.728, \quad \hat{p} \approx 0.233,$$

so denial probability drops by about 9.2 percentage points.

For a White applicant at same ratio change, probability drop is smaller because nonlinear marginal effect depends on index location.

12.2 Example B: O-Ring Failure and Binary Models

LPM estimate:

$$\widehat{OFail} = 2.858 - 0.037 \text{Temperature}.$$

At 31 degrees Fahrenheit,

$$\widehat{OFail} = 2.858 - 0.037(31) \approx 1.71,$$

which is outside $[0, 1]$, highlighting LPM limitations for binary outcomes.

Logit estimate:

$$\mathbb{P}(OFail = 1 \mid T) = \Lambda(15.297 - 0.236T).$$

Change from 80 to 70 degrees gives large increase in failure probability, and change from 60 to 50 degrees gives a different increase; effect is not constant because logit is nonlinear.

12.3 Example C: Instrument Validity for Returns to Education

In wage equation with endogenous education,

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + W'\gamma + u,$$

candidate instruments must satisfy relevance and exogeneity.

Discussion template:

- Parent education: often relevant, exogeneity debated through family background channels.
- Region indicators: often relevant, exogeneity depends on whether region has direct wage effects beyond controls.
- Test scores: typically correlated with ability in u , often invalid as exclusion restrictions.

12.4 Example D: Two-Group Restaurant DiD

Given means:

Low wage: $19.56 \rightarrow 20.88$ ($\Delta = +1.32$),

High wage: $22.25 \rightarrow 20.21$ ($\Delta = -2.04$).

DiD estimate:

$$\hat{\Delta}_{DiD} = 1.32 - (-2.04) = 3.36.$$

If $SE(\hat{\Delta}) = 1.48$, then

$$t = 3.36/1.48 \approx 2.27,$$

which rejects zero at conventional two-sided 5 percent level.

12.5 Example E: Cigarette Tax Natural Experiment

Regression in change form:

$$\Delta CigCons_i = \beta_0 + \beta_1 TaxIncrease_i + u_i.$$

Estimated $\hat{\beta}_1 < 0$ is directionally consistent with demand logic. Precision depends strongly on between-state heterogeneity and limited sample size.

12.6 Example F: Panel FE and Policy Effects

With state-year data,

$$Fatality_{st} = \alpha_s + \lambda_t + \beta BeerTax_{st} + u_{st}.$$

Entity FE absorb time-invariant state traits; time FE absorb common shocks. Inference should cluster by state to address serial correlation in u_{st} .

12.7 Example G: Counting FE Parameters

Suppose panel with 9 regions and 7 years, one slope regressor, and intercept. Two-way FE coefficient count is

$$1 \text{ (intercept)} + 1 \text{ (slope)} + (9 - 1) + (7 - 1) = 16.$$

This count is useful for checking software output and ensuring no dummy-variable trap.

12.8 Example H: Region FE Restriction Test

Replacing 49 state effects with 8 region effects imposes linear restrictions that states in a region share the same fixed effect. This can be tested with an F-test comparing unrestricted and restricted FE specifications. Rejecting indicates meaningful heterogeneity inside regions.

12.9 Example I: Weak-IV Sensitivity

Suppose OLS estimate of education return is 0.10 and IV estimate is 0.04. If first-stage F is 3, the IV estimate may still suffer weak-identification distortion. Interpretation should emphasize design uncertainty and report weak-IV robust inference, not only conventional t statistics.

12.10 Example J: Difference Between FE and Pooled OLS

In a country-growth panel, pooled OLS may show strong positive association between democracy and income. Adding country and year fixed effects can shrink estimate sharply because time-invariant institutions and global shocks are absorbed. This change is evidence that between-country confounding was substantial.

13 Extended Derivation Workbook

13.1 Deriving IV from Orthogonality Conditions

Moment condition under valid instrument:

$$\mathbb{E}[Z_i(Y_i - \beta_0 - \beta_1 X_i)] = 0.$$

After demeaning (or adding intercept separately),

$$\mathbb{E}[(Z_i - \mathbb{E}[Z_i])(Y_i - \beta_1 X_i)] = 0,$$

which implies

$$\beta_1 = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)}.$$

Sample analog yields just-identified IV estimator.

13.2 Generated Regressor View of 2SLS

First stage:

$$X_1 = P_Z X_1 + M_Z X_1 = \hat{X}_1 + \tilde{v}, \quad M_Z = I - P_Z.$$

Second stage uses \hat{X}_1 , which keeps variation in endogenous regressor explainable by instruments and exogenous controls.

13.3 Bias Direction Under Weak Instruments

With one endogenous regressor and weak first stage, finite-sample 2SLS can lie between OLS and true parameter. If endogeneity pushes OLS upward and true effect is lower, weak-IV 2SLS often remains upward-biased.

13.4 Within Estimator as OLS on Residualized Data

Let Q be residual maker that partials out entity dummies. Then FE estimator is

$$\hat{\beta}^{FE} = (X'QX)^{-1} X'Qy.$$

This is exactly OLS after removing entity means from both X and y .

13.5 Two-Way FE Residualization

With entity dummies D and time dummies F , partial out both sets by projection. Operationally, software handles this by absorbing fixed effects; algebraically it remains a projection estimator.

13.6 DiD in First-Difference Form

For two periods $t = 0, 1$:

$$\Delta Y_i = Y_{i1} - Y_{i0} = \gamma_0 + \gamma_1 D_i + \eta_i.$$

Here γ_1 equals DiD estimand, so 2x2 DiD can be estimated as a cross-sectional regression on outcome differences.

13.7 Standard Error Choice in DiD

If treatment varies by group and group outcomes are serially correlated, cluster by group (not by group-time cell) for valid inference. Ignoring this often understates uncertainty.

13.8 Linear Restrictions in FE Panels

Suppose we estimate

$$Y_{it} = \alpha_i + \lambda_t + \beta_1 X_{1,it} + \beta_2 X_{2,it} + u_{it}$$

and want to test $H_0 : \beta_1 + \beta_2 = 0$. Define $c = (0, 0, 1, 1)'$ over the slope block and compute

$$t = \frac{c' \hat{\beta}}{\sqrt{c' \widehat{\text{Var}}(\hat{\beta}) c}},$$

using cluster-robust covariance when clustering is appropriate.

13.9 How Unclustered SE Can Overstate Precision

If $u_{it} = a_i + \varepsilon_{it}$ with persistent entity component a_i , then observations within entity are correlated. Treating them as independent effectively counts repeated information multiple times and mechanically shrinks estimated standard errors.

13.10 DiD Decomposition with Covariates

In

$$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \cdot Post_t) + W_{it}' \gamma + u_{it},$$

the interaction coefficient still captures differential pre-post change, but now conditional on W_{it} . Thus interpretation is “DiD net of observed covariate shifts”.

13.11 ATT Interpretation in DiD

Under parallel trends and no anticipation, 2x2 DiD identifies average treatment effect on the treated group in the post period. This is often the policy-relevant estimand and should be labeled explicitly in writeups.

14 Assumption and Interpretation Checklist

14.1 Binary Outcome Models

1. Is the dependent variable truly binary?
2. If using LPM, are robust SE reported?

3. Are predicted probabilities and marginal effects interpreted correctly?
4. For nonlinear models, is marginal-effect evaluation point stated (MEM vs AME)?

14.2 IV Design

1. Relevance: is first stage strong enough?
2. Exogeneity: is exclusion restriction economically defensible?
3. If overidentified, do diagnostics support instrument set?
4. Do conclusions change materially between OLS and IV in expected direction?

14.3 Panel and DiD

1. Are entity and time effects aligned with the design?
2. Are standard errors clustered at the right level?
3. Is parallel trends argument supported by design and pre-treatment data?
4. Are treatment timing and group definitions explicit?

15 Common Failure Modes

- treating a weak instrument as valid identification,
- interpreting nonlinear-model coefficients as constant probability effects,
- forgetting that FE removes time-invariant regressors,
- using heteroskedastic robust SE when cluster SE are needed,
- reporting DiD estimate without discussing parallel trends.

16 Proof Appendix

16.1 LPM Conditional Variance Identity

For binary Y with $p = \mathbb{P}(Y = 1 \mid X)$:

$$\text{Var}(Y \mid X) = \mathbb{E}[Y^2 \mid X] - \mathbb{E}[Y \mid X]^2 = p - p^2 = p(1 - p).$$

Since $u = Y - p$, this is also $\text{Var}(u \mid X)$.

16.2 IV Orthogonality Proof

Valid instrument implies

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

With structural equation,

$$\mathbb{E}[Z_i Y_i] = \beta_0 \mathbb{E}[Z_i] + \beta_1 \mathbb{E}[Z_i X_i].$$

After removing intercept term (demeaning or adding constant instrument), solve for β_1 as ratio of covariances.

16.3 FWL Proof Sketch for FE

Regressing y on (X, D) where D are entity dummies gives slope

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

with $M_D = I - D(D'D)^{-1}D'$. M_D performs within-entity demeaning, proving LSDV and within estimators are equivalent.

16.4 DiD Coefficient Identity

In

$$Y = \beta_0 + \beta_1 D + \beta_2 Post + \beta_3 (D \cdot Post) + u,$$

cell means map one-to-one to parameter sums. Taking treated post-pre minus control post-pre leaves only β_3 .

16.5 Cluster-Robust Intuition

Classical variance uses $\sum_i x_i x_i' \hat{u}_i^2$ and treats observations as independent. Cluster variance replaces this with block sums $X_g' \hat{u}_g \hat{u}_g' X_g$, keeping within-cluster covariance terms.

16.6 FE and FD Equivalence When T Equals Two

With two periods,

$$\bar{Y}_i = \frac{Y_{i1} + Y_{i2}}{2}, \quad Y_{i2} - \bar{Y}_i = \frac{1}{2}(Y_{i2} - Y_{i1}),$$

and similarly for X and u . So FE transformed equation is a constant multiple of FD equation, implying identical slope estimates.

16.7 LATE Decomposition Identity

For binary instrument under monotonicity,

$$\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0] = \mathbb{P}(\text{complier}),$$

and

$$\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0] = \mathbb{P}(\text{complier}) \cdot \mathbb{E}[Y(1) - Y(0) | \text{complier}].$$

Dividing gives Wald equals LATE.

16.8 Serial Correlation and DiD Variance

Let treatment vary at group level and errors follow

$$u_{gt} = \rho u_{g,t-1} + \eta_{gt},$$

with $|\rho|$ near one. Then adjacent period residuals in a group are highly correlated; naive iid formulas understate uncertainty of group-level treatment effects. Clustering by group restores valid large-sample variance under cross-group independence.

17 Practice Set with Solution Sketches

17.1 Q1: LPM Heteroskedasticity

Show in two lines why LPM error variance depends on X .

Sketch: use $u = Y - \mathbb{E}[Y | X]$, binary identity $Y^2 = Y$.

17.2 Q2: Nonlinear Marginal Effects

Why can two observations with identical X_j coefficients have different marginal effects in probit/logit?

Sketch: derivatives are coefficient times a density term that depends on full index $X'\beta$.

17.3 Q3: OLS Endogeneity Bias

Derive $\text{plim } \hat{\beta}_1^{OLS}$ when $\text{Cov}(X, u) \neq 0$.

17.4 Q4: IV Consistency

State and prove the two key IV conditions in covariance form.

Sketch: relevance and exogeneity map directly into covariance ratio identity.

17.5 Q5: First Stage Strength

Why does low first-stage F threaten valid t-tests in second stage?

Sketch: weak relevance yields non-normal finite-sample behavior and severe size distortion.

17.6 Q6: Overidentification

What does rejection in an overidentification test imply?

Sketch: at least one exclusion restriction is inconsistent with data and model assumptions.

17.7 Q7: FE Identification

Why is a time-invariant regressor unidentified in entity FE?

Sketch: within transformation subtracts entity mean, making regressor identically zero.

17.8 Q8: Cluster SE

Give one reason state-level panel regressions usually cluster by state.

Sketch: shocks persist within state over time, violating independence of within-state errors.

17.9 Q9: DiD Estimand

Write the 2x2 DiD estimand from group-time means and show how it maps to β_3 .

17.10 Q10: Parallel Trends

Why is parallel trends about untreated potential outcomes, not observed outcomes?

Sketch: observed treated outcomes include treatment effects after intervention.

17.11 Q11: Repeated Cross Section DiD

Can DiD be estimated if units differ across periods?

Sketch: yes, if group definitions are stable and assumptions hold at group-mean level.

17.12 Q12: Binary Model Choice

When might LPM still be preferred over probit/logit?

Sketch: transparent marginal effects, easy FE/IV extensions, and robust SE handling.

17.13 Q13: Policy Interpretation

In log-level model, what is exact percent effect of one-unit policy change?

Sketch: $100 \times (e^\beta - 1)$.

17.14 Q14: DiD Standard Errors

What common mistake causes overconfident DiD inference?

Sketch: not clustering at treatment-assignment group level.

17.15 Q15: Instrument Screening

List two instruments that are likely invalid for education in wage equation and explain why.

Sketch: variables directly linked to ability/wage determinants violate exclusion.

17.16 Q16: FE vs Pooled OLS

Why can FE and pooled OLS slopes differ dramatically in panel policy studies?

Sketch: pooled OLS mixes cross-sectional confounding with within-unit policy variation.

17.17 Q17: LATE

Why does the Wald ratio identify complier effects rather than full-population ATE?

Sketch: instrument shifts treatment only for marginal units whose treatment status responds to Z .

17.18 Q18: Monotonicity

What is the economic meaning of the no-defiers assumption?

Sketch: no unit moves opposite treatment direction when encouragement increases.

17.19 Q19: FE vs RE Choice

State one condition under which random effects is consistent and one reason it may fail in practice.

Sketch: consistency needs $\text{Cov}(\alpha_i, X_{it}) = 0$; omitted persistent heterogeneity often violates this.

17.20 Q20: Dynamic Panel

Why does including lagged dependent variable in short-panel FE create bias?

Sketch: transformed lagged outcome correlates with transformed error term.

17.21 Q21: Event Study

What do statistically significant lead coefficients imply in an event-study DiD plot?

Sketch: evidence against common untreated trends or presence of anticipation effects.

17.22 Q22: DiD With Controls

How does adding covariates change interpretation of interaction coefficient in DiD regression?

Sketch: treatment effect is differential change conditional on included observables.

17.23 Q23: Overidentification

If overidentification test rejects, can you still trust coefficient magnitude?

Sketch: not without additional justification; rejection signals at least one invalid instrument.

17.24 Q24: Cluster Level

In a classroom-level intervention with school-level policy rollout, where should clustering usually occur?

Sketch: at school level, where treatment and correlated shocks are assigned.

18 Derivation and Proof Summary

1. Binary outcome models estimate conditional probabilities; nonlinear links enforce bounds.
2. IV isolates exogenous variation in endogenous regressors through relevance and exogeneity.
3. 2SLS is projection-based IV in matrix form and requires strong first stages for reliable inference.
4. Fixed effects remove time-invariant confounding via within-unit variation.
5. Clustered SE are usually required in panel policy settings.
6. DiD identifies treatment effects from differential pre-post changes under parallel trends.

19 Bridge to Integrated ECON 120 Notes

These 120C tools complete the sequence started in 120A and 120B: distributional foundations, regression inference, and causal designs for observational data.