639 Final Reference Sheet

OLS Assumptions:

1) The conditional distribution of u_i given X_i has a mean of 0. The formal statement is: $E(u_i|X_i) = 0$, which implies that X_i and u_i are **uncorrelated**.
2) $(X_i, Y_i), i = 1, ..., n$ are independently and identically distributed (i.i.d). If X_i and Y_i are drawn from the same population, they will have the same distribution, and if they are drawn randomly, then the selection of any X_i or Y_i into the sample should be independent.
3) Large outliers are unlikely –> Large outliers can make the model results misleading!
4) Errors are homoskedastic – In practice just use robust standard errors, which adjust for homoskedasticity.

Population Regression line (no hats!)

$$Y_i = \beta_0 + \beta_1 X_1 i + \beta_2 X_2 i + \ ... \ \beta_k 1 X_k i + u_i, i = 1, \ ... \ , n$$

Sample Regression/OLS Regression line

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1} X_i, \text{ i = 1, ... } , n$$

Omitted Variable Bias

OVB occurs when two criteria are met:
1) the omitted variable (X_2) is correlated with the included regressor (X_1).
2) the omitted variable (X_2) is a determinant of the dependent variable (Y) Note: Omitted variable bias means that the first least squares assumption for causal inference – that $E(u_i|X_i) = 0$, **does not hold**, resulting in a biased estimator. This cannot be fixed with large samples!
Use this table: where α_1 is the bivariate model regressor, and β_1 is original regressor after adding second variable, and γ is the impact of β_2 on β_1

		How is X_2 related to Y?		
$Corr(X_1, X_2)$		$\beta_2 < 0$	$\beta_2 = 0$	$\beta_2 > 0$
Or: How is X_1	$\gamma < 0$	Positive Bias ($\alpha_1 > \beta_1$)	No Bias	Negative Bias ($\alpha_1 < \beta_1$)
related to X_2	$\gamma = 0$	No Bias	No Bias	No Bias
	$\gamma > 0$	Negative Bias ($\alpha_1 < \beta_1$)	No Bias	Positive Bias ($\alpha_1 > \beta_1$)

Non-linearities

Transformation	Model	Interpretation
No Transformation	$Y = \beta_0 + \beta_1 X$	A 1 unit increase in X is associated with an average change of β_1 units in Y
Binary X Variables	$Y = \beta_0 + \beta_1 X$	A 1 unit increase in X is associated with an average change of β_1 units in Y
Log-transformed predictor (level-log)	$Y = \beta_0 + \beta_1 log(x)$	A 1% change in X is associated with an average change of β_1 /100 units in Y
Log-transformed outcome (log-level)	$log(Y) = \beta_0 + \beta_1 X$	A 1 unit increase in X is associated with an average change of $100 \cdot \beta_1\%$ in Y
Log-log model	$log(Y) = \beta_0 + \beta_1 log(X)$	A 1% increase in X is associated with a β_1 % change in Y
Quadratic	$Y = \beta_0 + \beta_1 X + \beta_2 X^2$	$\Delta Y = (\beta_1 + 2\beta_2 X)$
Linear Prob. Model (LPM) ($Y \in \{0, 1\}$)	$Pr(Y = 1) = \beta_0 + \beta_1 X$	A 1 unit increase in X is associated with a β_1 change in the probability of Y
Logit	$Pr(Y = 1) = \frac{1}{1+e^{-\beta_0 + \beta_1 X}}$	A 1 unit change in X is associated with a β_1 change in the log-odds of success:failure in Y
Probit	$Pr(Y = 1|X) = \phi(\beta_0 + \beta_1 X)$	A 1 unit increase in X is associated with an β_1 increase in **z-score**

Quadratic note: Exact Method

$$(\hat{Y_1}|X_1 = x) + \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$$
$$(\hat{Y_2}|X_2 = x) + \beta_0 + \beta_1 X_2 + \beta_2 X_2^2$$
$$\Delta Y = \hat{Y_2} - \hat{Y_1}$$

Logit Note: If coefficient estimate is > .10, YOU MUST EXPONENTIATE. Interpreting as a percent is not precise!!!!!

Internal Validity	External Validity
Did we estimate an unbiased causal effect for our sample? Fails when treatment and control groups are different in ways (beside the treatment) that may affect the outcome of interest	Can we extrapolate these estimates to other populations? Fails when the treatment effect is different outside the evaluation

Internal Validity	External Validity
Threats: Non-compliance, attrition, evaluation-driven effects	

Interaction Terms

- Binary/Indicator Variables: Interpret regression results by multiplying all the category coefficients by 0. That is the estimate for the reference group, and all coefficients are interpreted relative to the reference group. To test, use the F-test. The null hypothesis (H_0) for the F-test is that *all* of the coefficients are zero.

Types of interactions:

- binary and continuous
- binary and binary
- continuous and continuous

Term	Notation	Interpretation
Indicator	$Y = \beta_0 + \beta_1 X + \beta_2 D$, where $D \in \{0, 1\}$	Allows for different intercepts: $D = 0 \rightarrow \beta_0$, $D = 1 \rightarrow \beta_0 + \beta_2$
Interaction	$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 XD$	Interaction between continuous variable X and binary variable D. Intuition: $Y = \beta_0 + \beta_2 D + (\beta_1 + \beta_3 D)X$, β_3 is the additional increase in the slope of X when D increases by 1 unit.

Fixed Effects

- **Entity (or State) FEs**: Control for unobservables that vary across entities but are fixed over time
- **Time FEs**: Control for unobservable variables that vary over time but are fixed across entities
- Control only for characteristics that do not vary within the group for which you have a fixed effect
  - e.g. district spending per student, student population per district
- Group fixed effects do not control for characteristics that vary between observations within groups
  - e.g. family income, age

Difference-in-Differences

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Treatment_i + \beta_3 Treatment_i x Post_t + \epsilon$$

Where:
- Y_it: the outcome for i at time t - $Post_i$: a dummy for all time periods **after** the treatment
- $Treatment_i$: a dummy variable for all entities that receive treatment
- $Treatment_i x Post_i$: an indicator for all treatment observations after the treatment has happened
- β_3 **is the estimated effect of the treatment**

Assumptions:
1. Common Trends: Treatment/intervention and control groups have Parallel Trends in outcome
2. Confounding factors: Other policies/changes that occur at the same time as the treatment that disproportionately impacted the treated sample

	Control	Treatment	Difference
Pre-	b0	b0 + b1	b0
Post	b0+b2	b0+b1+b2+b3	b1+b3
Difference	b2	b2+b3	b3 (Effects!)

Regression Discontinuity

Assumptions:
1. Cannot manipulate threshold
2. No other policy is happening at the cutoff

$$Y = \beta_0 + \beta_1 Treatment + \beta_2 RunningVar + \beta_3 Treat * RunningVar + u$$

Where Treatment is determined by a *score* or **threshold** or **running variable** (eligibility):
- Simple version – Regress outcome on treatment and function of score/running variable: $Y = \beta\text{Treatment} + f(\text{score}) + \epsilon$
- Better – Allow function to vary between treated and non-treated units: $Y = \beta\text{Treatment} + f_1(\text{score}) \cdot (\text{score} < 0) + f_2(\text{score}) \cdot (\text{score} \geq 0) + \epsilon$
Where $f(\text{score})$, $f_1(\text{score})$, and $f_2(\text{score})$ are smooth functions below and above the cutoff, and β is the discontinuous jump in Y at the cutoff.

Two types:
- Sharp: Treatment probability goes from $0 \rightarrow 1$ as running variable crosses the threshold.
- Fuzzy: Treatment probability increases discontinuously (from some level greater than 0 to some level less than 1) as the assignment/running variable crosses the threshold.
- α: Regress outcome on score, THEN
- δ: Regress treatment on score
- **Effect of treatment on Y**: $\beta = \frac{\alpha}{\delta}$ (initial effect rescaled by probability of compliance)