APPLIED ECONOMETRICS SUMMARY

Lec 1. Potential outcomes framework

- SDs in {brackets}; SEs in (parentheses)
- Though reject null (two groups different), not causal
- **Causal/Treatment Effect**: $Y_{1i} - Y_{0i}$
  ATE = $Avg_n[Y_{1i} - Y_{0i}] = \frac{1}{n}\Sigma_{i=1}^{n} Y_{1i} - \frac{1}{n}\Sigma_{i=1}^{n} Y_{0i}$
- Constant-effects assumption: $Y_{1i} = Y_{0i} + \kappa$
- Δgroup mean = $Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$
  $= \kappa + Avg_n[Y_{0i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$
  Δgroup mean = Treatment Effect + Selection Bias
- Randomized trial can eliminate selection bias
  $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$
- With sufficiently n, the Law of Large Number insures that conditional averages are close to conditional expectations, reveals $\kappa + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = \kappa$

Lec 2. Randomized Trials

- E(Y): fixed feature of population Y $\rightarrow$ parameter
  For samples from pop, many possible $\bar{Y} \equiv Avg_n[Y_i]$
  $\bar{Y}$: unbiased estimator of the pop mean, $E[\bar{Y}] = E[Y_i]$
- Variability: $V(Y_i) = \sigma_Y^2 = E[(Y_i - E[Y_i])^2]$, parameter $S(Y_i)^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2$, unbiased estimator of σ_Y^2
- $\bar{Y}$ is a random variable;
  Replace σ with $S(Y_i)$: $\hat{SE}(\bar{Y}) = \frac{S(Y_i)}{\sqrt{n}}$
- CLT: N sufficiently large, sampling distribution $Y_i \sim N$

$$t(\mu) = \frac{\bar{Y} - \mu}{\hat{SE}(\bar{Y})}$$

- We reject the null at the $\alpha = 0.05$ level if $|t| > 1.96$
  We reject the null at the $\alpha = 0.01$ level it $|t| > 2.58$
  We reject the null at the $\alpha = 0.001$ level it $|t| > 3.29$
- **Difference between group means**

$$t(\mu) = \frac{\bar{Y}^1 - \bar{Y}^0 - \mu}{\hat{SE}(\bar{Y}^1 - \bar{Y}^0)}$$

$$\hat{SE}(\bar{Y}^1 - \bar{Y}^0) = S(Y_i)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $S(Y_i)$ is pooled sample standard deviation
$S(Y_i) = \sqrt{pooled\_prop \cdot (1 - pooled\_prop)}$
$S(Y_i) = \sqrt{\frac{n_1 * s_1^2 + n_2 * s_2^2}{n_1 + n2}}$

Lec 3. Experimental Design

- Type I error (α) - reject null when null is correct
- Type II error (β) - fail to reject null when null is false
- **Power** of a test = $1 - \beta$: probability of making the correct decision if the alternative hypothesis is true
- A test is under-powered if it has power < 0.8
- **Increase power**:
  1. as sample size increases: larger sample sizes provide more information and reduce variability in estimates
  2. as effect size increases: H0 and H1 are further away, easier to distinguish
  3. as α increase, while increasing Type I errors

Intraclass correlation (ICC):

- Subjects may have related outcomes, not independent
- $0 < \rho < 1$, $\rho = 0.1$ (plausible value from literature)

$$\rho = \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_e^2}$$

where σ_γ^2 is variance of γ_j, across classrooms

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \rightarrow \sqrt{\frac{1 + (n_j - 1) * \rho}{n_1} + \frac{1 + (n_j - 1) * \rho}{n_2}}$$

- If we assume independence and does not accounts for ICC, underestimate variance of mean, larger t-score, reject more often
- *Clustering*: grouping observations into clusters or groups based on certain characteristics to account for ICC and heteroskedasticity

Lec 4. Regression

- Experiments often impossible $\Rightarrow$ regression framework

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i$$

- A_i or X_i: control variables, a set of dummy variables that identify individual's application and acceptance set
- *Fitted value*: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} P_i + \hat{\gamma} A_i$
- *Estimated residuals*: $Y_i - \hat{Y}_i$
- β is a weighted sum of the within-group differences
- OLS: $\hat{\alpha}$ and $\hat{\beta}$ are chosen to minimize RSS: $\Sigma_{i=1}^{n} e_i^2$
- Conditional expectation function (CEF): collection of conditional expectations $E[Y_i|X_i = x]$ over all possible values of X_i
- Regression is estimating the parameters in a linear CEF

$$E[\ln Y_i | P_i, GROUP_i, SAT_i, \ln PI_i]$$

$$= \alpha + \beta P_i + \Sigma_{j=1}^{150} \gamma_j GROUP_{ji} + \delta_1 SAT_i + \delta_2 \ln PI_i$$

- **Three steps of OLS**:
  1. matches subjects by value of covariates
  2. compares the outcome between treatment and control group of matched subjects for each possible combination of the conditioning variables
  3. produces a single average by averaging all of these cell-specific contrasts

$$\beta = E[\ln Y_i | P_i = 1, ...] - E[\ln Y_i | P_i = 0, ...]$$

- Any weighted average of cell-specific estimates will be an unbiased estimate of β
- **Why adding covariates/controls**: matching subjects into groups on patterns to ensure that subjects are similar in terms of ... characteristics that affect ... If subject within groups are indeed comparable, we have taken care of selection bias
- We may exclude controls if **randomized experiment**, no longer have to worry about selection bias, unbiased

Lec 5. Regression Analysis

- OLS estimator is the sample mean

$$Y_i = \alpha + \beta X_i + e_i$$
$$\hat{\beta} = \frac{Cov(X_i, Y_i)}{Var(X_i)} = \frac{\text{sample covariance}}{\text{sample variance of } X_i}$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

- $E[e_i] = 0$
- $E[X_i e_i] = 0, E[e_i|X_i] = E[e_i]$, residuals are part of outcome Y_i uncorrelated with the regressors
- For the k-th regressor:

$$\hat{\beta}_k = \frac{Cov(\tilde{X}_{ki}, Y_i)}{Var(\tilde{X}_{ki})}$$

where $\tilde{X}_{ki}$ is residual from the regression of X_{ki} on all other k-1 regressors; the part of X_{ki} that is not correlated with the other regressors; **paritial effect** of X_1 on Y that cannot be explained by other regressors

$$\ln Y_i = \alpha^l + \beta^l P_i + \gamma A_i + e_i^l$$
$$\ln Y_i = \alpha^s + \beta^s P_i + e_i^s$$

Omitted Variable Bias (OVB):

$$\beta^s = \beta^l + \frac{Cov(A_i, P_i)}{V(P_i)}\gamma = \beta^l + \pi_i\gamma$$

- short equals long plus the effect of omitted times the regression of omitted on included
- If all relevant regressors included, OLS can estimate treatment effects

Lec 6. Regression Analysis Application I

- SE of estimated $\hat{\beta}$ in a **bivariate regression**:

$$SE(\hat{\beta}) = \frac{\sigma_e}{\sqrt{n}} \times \frac{1}{\sigma_X}$$

- **Smaller SE** (*ceteris paribus*) when:
  1. larger samples (n large)
  2. regression line fits well (σ_e small)
  3. X is spread out (σ_X large)
- SE of estimated $\hat{\beta}_k$ from a **multiple regression**:

$$SE(\hat{\beta}_k) = \frac{\sigma_e}{\sqrt{n}} \times \frac{1}{\sigma_{\tilde{X}_k}}$$

- Above assumes homoskedasticity: variance of residuals is unrelated to regressors
- Allow for heteroskedasticity $\rightarrow$ robust standard errors: variance of residuals is varying across observations
- **Coefficient of determination** (R^2): $0 < R^2 < 1$
- R^2 is just a useful heuristic

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- $R^2 \uparrow$, SSR$\downarrow$, standard errors for regression coefficient$\downarrow$
- Adjusted R^2
- **Heteroskedasticity** only affect estimated SE but not coefficients, higher SE, t-score smaller, rejects less often

Diamond Application:

- as.factor() create one dummy for each category
- Interpretation of β_1: Controlling for clarity, on average, a one *carat* increase in weight is associated with an increase in the value of the diamond by \$12,350, holding all else constant
- Interpretation of R_2: approximately 96% of the variation in the dependent variable is explained by the model
- Using hetero (`hetero, HC1`), robust standard errors account for possible heteroskedasticity of the error term; SE on *carat* is larger than the ordinary SE
- Interpretation of dummies: Controlling for *carat*, being in category 3 on average is associated with an increase in the value of the diamond by \$4,253, compared to the reference group category 2
- No meaningful interpretation of intercpet since no diamond with *carat* = 0
- Difference between groups: corresponding $\Delta\beta$
- OVB: Since *carat* and *clarity* are negatively correlated and clarity is positively correlated with *price*, omitting *clarity* will lead to a negative bias, underestimating the coefficient on *carat*; R^2 drops

Interpretation of coefficients

Model	Dependent	Independent	β_1 interpretation
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$log(x)$	$\Delta y = (\beta_1/100)(\%\Delta x)$
Log-level	$log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$log(y)$	$log(x)$	$\%\Delta y = \beta_1(\%\Delta x)$
Quadratic	y	$x + x^2$	$\Delta y = (\beta_1 + 2\beta_2 x)\Delta x$

- *Log-level*: On average, each additional unit in X_i is **associated with** a $100\beta_1\%$ increase in Y, holding all else constant
- **controlling for** <control>, ...
- **on average**, ...
- ..., **holding all else constant**

Other

- **Selection bias**: the difference in potential outcomes between the group that is selected (self-selected) into treatment and the group not selected into treatment.
- An example of selection bias: Measure the return to grad by solely comparing the average earnings between people that went to grad school and ones that did not. Individuals that self-select themselves into grad school may have different characteristics (high skills, high aspirations, previous network) that would benefit them regardless of graduate education. Overestimating the casual effect of grad education on earnings
- In a **randomized control** trial, we expect treatment and control group to have same characteristics on average, *balancing test*: check if covariates (control variables) are statistically equal across treatment and control group
- Use sample sizes [100,100] instead of [150, 50] for a smaller SE:

$$Var(\bar{Y}^1 - \bar{Y}^0) = \sigma_Y^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

Lec 7: Regression Analysis: Application II

- **Unbiased estimation** $\Rightarrow$ On average, treatment and control groups are balanced in both observed and unobserved characteristics. When randomization is successful, the groups should be comparable, any differences in outcomes can be attributed to the treatment effect
- **Covariates** = control variables
- If randomized into groups, no need to include covariates
- Adding statistically significant controls will improve SE since variance of $e_i \downarrow$ and $R^2 \uparrow$, increase efficiency

Lec 8: Regression Analysis: Advanced Modeling

- **Fixed effects**: dummies that account for unobserved individual heterogeneity that do not vary over time
- **State-by-month fixed effects** (α_{sm}): capture the unobserved, time-invariant characteristics specific to each state for each month
  - state-specific policies, culture

Instrumental Variables

- **Why IV**: randomized trials expensive and regression inadequate, omitted variable bias unavoidable
- Instrument $Z_i \rightarrow$ Treatment $D_i \rightarrow$ Outcome Y_i
- **Good IV Requirements**:
  1) *Relevance condition*: instrument Z_i has a causal effect on treatment D_i.
  2) *Independent assumption*: instrument Z_i is randomly assigned or "as good as randomly assigned," being unrelated to the omitted variables (exogenous).
  3) *Exclusion restriction*: single channel through which instrument Z_i affects outcomes Y_i, which is through observed differences in D_i.
  4) *Monotonicity*
  First Stage (Treatment D_i on instrument Z_i):

$$D_i = \alpha_1 + \phi Z_i + e_{1i}$$

$$\phi = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

  Reduced Form (Outcome Y_i on instrument Z_i):

$$Y_i = \alpha_0 + \rho Z_i + e_{0i}$$

$$\rho = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

- Casual effect of interest **(LATE)**:

$$\lambda = \frac{\text{Reduced Form}}{\text{First Stage}}$$

$$\lambda = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

$$\lambda = \frac{\rho}{\phi}$$

- **Never-takers**: $E[Y_{0i}|N_i = 1]$, regardless of the treatment, would NEVER take the treatment;
- **Always-takers**: $E[Y_{1i}|A_i = 1]$, regardless of being in the control, would ALWAYS take the treatment;
- **Compliers**: $E[Y_{1i}|C_i = 1]$, COMPLY: take the treatment if assigned to treatment group and do not take the treatment if assigned to control group.
- **Monotonicity assumption**: no defiers; instrument affects the treatment in one direction only
- **LATE Theorem**: for any randomly assigned instrument with a nonzero first stage, satisfying both monotonicity and an exclusion restriction, the ratio of the reduced form to first stage is LATE, on **compliers**

$$\lambda = \frac{\rho}{\phi} = E[Y_{1i} - Y_{0i}|C_i = 1]$$

  We can only measure impact of treatment on people who respond to the instrument; unsure about how results generalize to other populations (**external validity**)
- **Two stage least squares (2SLS)**, allows more general models: more than one instrument, including covariates
  First Stage (Treatment D_i on instrument Z_i):

$$D_i = \alpha_1 + \phi Z_i + e_{1i} \text{ and forms } \textit{fitted } \hat{D}_i$$

- Fitted $\hat{D}_i$ includes variation explained by chosen Z_i
  Second Stage (Outcome Y_i on treatment $\hat{D}_i$):

$$Y_i = \alpha_2 + \lambda \hat{D}_i + e_{2i}$$

$$\lambda_{2SLS} = \frac{\rho}{\phi} = \lambda$$

- Unlike OLS, 2SLS is subject to *finite* sample bias
- **F statistic** tests the joint hypothesis that coefficients of all instrumental variables are equal to zero
- Check if F-statitic > 10 in the first stage
- F-statitic = t^2 for one instrument

Lec 11: IV Applications: Class Size

- 1) Use OLS table, see OVB, thus use IV;
  2) Think of IV, requirements, formulate First stage, Reduced form and Second stage;
  3) Use First stage and Reduced form coefficient from table for IV estimates;
  4) Use 2SLS for IV estimates and standard errors;
  5) Check first stage F-stat = t^2 > 10 for one instrument;
  6) Only Local Average Treatment Effect, not for every ...
- **Internal validity**: when the estimation strategy successfully uncovers a causal effect
- **External validity**: when those estimates are predictive of outcomes in other scenarios
- IV estimate is a local average treatment effect (LATE)
- IV $\rightarrow$ internal validity, not necessarily external validity
  We aren't learning the effect of D_i on Y_i for observations where Z_i doesn't explain D_i.
- Possible test for external validity problems: small RSS from first stage means Instrument Z_i explains well D_i

Other

- **Reverse causality** $\Rightarrow$ X explains Y and Y explains X.
- There could be intra-class correlation between MSAs; A large and populated MSA will have more IH lanes and smaller MSAs nearby can benefit from these interstate highways; cluster by MSA to reduce ICC
- Adding fixed effects helps to control for the unobserved heterogeneity and can provide more reliable estimates

```
# Demand: IV-2SLS Estimation
# feols(y ~ x_exog | x_endo ~ x_inst, data)
model_demand_iv <- feols(ln_q ~ ln_pbeef + ln_y + pop + cpi
| ln_pchick ~ ln_pf, data = data)

# ivreg(y ~ exo | endo | instrument)
model_demand_iv <- ivreg(ln_q ~ ln_pbeef + ln_y + pop + cpi
| ln_pchick | ln_pf, data = data)
```