Summary of the Econometrics II course

Tutorial 7

Stanislav Avdeev

Goal for today's tutorial

- Discuss the full course
  - Lecture 1: Binary choice models, censoring, truncation, and selection models (3-9)
  - Lecture 2: IV (10-13)
  - Lecture 3: Panel data models (14-20)
  - Lecture 4: Potential outcomes model (21-24)
  - Lecture 5: LATE and power analysis (25-27)
  - Lecture 6: DiD (28-30)
  - Lecture 7: RDD and RKD (31-34)

Lecture 1: Binary choice models

- Y_i can only take the values 0 or 1

$$Y_i = \begin{cases} 1 & \text{with probability } p_i \\ 0 & \text{with probability } (1 - p_i) \end{cases}$$

- For a binary model, the cumulative distribution function (cdf) is

$$p_i = P(Y_i = 1 | X_i) = F(X_i' \beta)$$

- For a binary model, the probability density function (density) is

$$f(Y_i | X_i) = p_i^{Y_i} (1 - p_i)^{1 - Y_i}$$

- To find β, use maximum likelihood function

$$\begin{aligned} L(\beta) &= \sum\nolimits_{i=1}^{N} \left[Y_i \ln p_i + (1 - Y_i) \ln(1 - p_i) \right] \\ &= \sum\nolimits_{i=1}^{N} \left[Y_i \ln F(X_i' \beta) + (1 - Y_i) \ln(1 - F(X_i' \beta)) \right] \end{aligned}$$

- Notice we did not specify a particular form of the cdf $F(X_i' \beta)$

Lecture 1: Binary choice models

- Linear probability model
  - $p_i = F(X_i'\beta) = X_i'\beta$
  - Marginal effect: $\frac{\partial p_i}{\partial X_{ik}} = \beta_k$
  - There is heteroskedasticity, so use robust s.e.
  - Estimated probabilities can be outside of the bounds
- Logit
  - $p_i = F(X_i'\beta) = \frac{exp(X_i'\beta)}{1+exp(X_i'\beta)}$ - the cdf of logistic distribution
  - Marginal effect: $\frac{\partial p_i}{\partial X_{ik}} = \frac{exp(X_i\beta)}{(1+exp(X_i\beta)^2}\beta_k$
  - MLE is not consistent if $F(\cdot)$ is incorrectly specified
- Probit
  - $p_i = F(X_i'\beta)$ = $\Phi(X_i'\beta)$ - the cdf of standard normal distribution
  - Marginal effect: $\frac{\partial p_i}{\partial X_{ik}} = \phi(X_i\beta)\beta_k$
  - MLE is not consistent if $F(\cdot)$ is incorrectly specified

Lecture 1: Latent structure

- Binary choice models are often written in terms of a latent structure with some latent (unobserved) variable

$$Y_i^* = X_i'\beta + U_i$$

- The observed outcome variable is

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0 \end{cases}$$

with

$$\begin{aligned} P(Y_i = 1|X_i) &= P(Y_i^* > 0|X_i) \\ &= P(X_i'\beta + U_i > 0|X_i) \\ &= P(-U_i < X_i'\beta|X_i) \\ &= F(X_i'\beta) \end{aligned}$$

where the cdf $F(\cdot)$ is symmetric

Lecture 1: Censoring and truncation

- The latent (unobserved) variable is

$$Y_i^* = X_i'\beta + U_i$$

- The observe outcome variable is

$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > c_i \\ c_i & \text{if } Y_i^* \leq c_i \end{cases}$$

- Censored observations are in the sample
  - for them $Y_i = c_i$ if $Y_i^* \leq c_i$
- Truncated observations are not in the sample
  - for them Y_i is missing if $Y_i^* \leq c_i$
- Ignoring censoring and truncation leads to a biased and inconsistent estimator

Lecture 1: Censoring and truncation

- To find θ, use maximum likelihood function. Assume $f^*(Y_i|X_i)$ is a density function of Y_i^*, then the cdf function of Y_i^* is

$$F^*(c_i|X_i) = P(Y_i^* < c_i|X_i) = \int_{-\infty}^{c_i} f^*(Y_i|X_i)dY_i$$

- Censoring
  - density function: $f(Y_i|X_i) = f^*(Y_i|X_i)^{d_i} F^*(c_i|X_i)^{1-d_i}$ with $d_i = 1$ for uncensored observations
  - log-likelihood function

$$L(\theta) = \sum_{i=1}^{N} \left[d_i \ln f^*(Y_i|X_i,\theta) + (1-d_i) \ln F^*(c_i|X_i,\theta)\right]$$

- Truncation
  - density function: $f(Y_i|X_i) = \frac{f^*(Y_i|X_i)}{P(Y_i^* > c_i)} = \frac{f^*(Y_i|X_i)}{1-F^*(c_i|X_i)}$
  - log-likelihood function

$$L(\theta) = \sum_{i=1}^{N} \left[\ln f^*(Y_i|X_i,\theta) - \ln(1-F^*(c_i|X_i,\theta))\right]$$

Lecture 1: Sample selection model

- The outcome variable is observed only for a selected sample
- The sample selection model has two stages
  - Selection equation

$$I_i^* = Z_i'\gamma + V_i$$

- The indicator function, based on I_i^*, takes two values

$$I_i = \begin{cases} 1 & \text{if } I_i^* > 0 \\ 0 & \text{if } I_i^* \leq 0 \end{cases}$$

  - Regression equation

$$Y_i^* = X_i'\beta + U_i$$

  - However, we observe only Y_i

$$Y_i = \begin{cases} Y_i^* & \text{if } I_i = 1 \\ \text{missing} & \text{if } I_i = 0 \end{cases}$$

Lecture 1: Sample selection model

- To estimate the sample selection model, we make an assumption that disturbances terms are bivariate normal

$$\begin{bmatrix} U_i \\ V_i \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right)$$

- Let us find expected value Y_i conditional on $I_i = 1$, i.e. observed Y_i

$$\begin{aligned} E[Y_i|I_i = 1, Z_i, X_i] &= E[X_i'\beta + U_i|I_i = 1, Z_i, X_i] \\ &= X_i'\beta + E[U_i|I_i = 1, Z_i, X_i] \\ &= X_i'\beta + E[U_i|Z_i'\gamma + V_i > 0, Z_i, X_i] \\ &= X_i'\beta + E[U_i| - V_i < Z_i'\gamma, Z_i, X_i] \\ &= X_i'\beta + \rho\sigma\frac{\phi(Z_i'\gamma)}{\Phi(Z_i'\gamma)} \end{aligned}$$

- If $\rho = 0$, i.e. if U_i and V_i are independent or when X_i and Z_i are uncorrelated, OLS estimator is consistent
- If $\rho \neq 0$, OLS estimator is inconsistent, and $\frac{\phi(Z_i'\gamma)}{\Phi(Z_i'\gamma)}$ is the Inverse Mills ratio which denotes selection bias

Lecture 2: IV

- If $E(U_i|X_i) \neq 0$, there is endogeneity problem
- In this case OLS provides a biased and inconsistent $\hat{\beta}$
- Sources of endogeneity
  - Omitted variables
  - Reverse causality
  - Measurement error
- A solution is to use an instrument that should be
  - Relevant: $\mathbf{cov}(Z_i, X_i) \neq 0$
  - Valid (exogenous): $\mathbf{cov}(Z_i, U_i) = 0$
- Use two-stage least squares (IV) estimator
  - First stage

$$X_i = \gamma_0 + \gamma_1 Z_i + V_i$$
$$\implies \hat{X}_i = \hat{\gamma}_0 + \hat{\gamma}_1 Z_i$$

  - Second stage

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + U_i^*$$
$$\implies \hat{\beta}_{1,2SLS}$$

Lecture 2: IV

- $\hat{\beta}_{1,2SLS}$ has the following form

$$\hat{\beta}_{1,\text{2SLS}} = \frac{\sum_{i=1}^{n}\left(Z_i - \bar{Z}_n\right)\left(Y_i - \bar{Y}_n\right)}{\sum_{j=1}^{n}\left(Z_j - \bar{Z}_n\right)\left(X_j - \bar{X}_n\right)}$$

- $\hat{\beta}_{1,2SLS}$ is consistent

$$\textbf{plim}_{n\to\infty}\hat{\beta}_{1,2SLS} = \frac{\textbf{cov}(Z_i, Y_i)}{\textbf{cov}(Z_i, X_i)} = \beta_1 + \frac{\textbf{cov}(Z_i, U_i)}{\textbf{cov}(Z_i, X_i)} = \beta_1$$

- $\hat{\beta}_{1,2SLS}$ is biased

$$E[\hat{\beta}_{1,2SLS}] = \beta_1 + \sum_{i=1}^{n}\mathrm{E}\left[\frac{\frac{1}{n}\left(Z_i - \bar{Z}_n\right)U_i}{\frac{1}{n}\sum_{j=1}^{n}\left(Z_j - \bar{Z}_n\right)\left(X_j - \bar{X}_n\right)}\right] \neq \beta_1$$

- Do you want to derive more consistency and unbiasedness of estimators? Take the core course Advanced Econometrics I

Lecture 2: IV

- To test validity, use the Sargan test (over-identification required)
  - H_0: all instruments are valid
  - Find the second-stage residuals and regress them on the instruments

$$U_i = \delta_0 + \delta_1 Z_{1,i}, + \ldots + \delta_M Z_{M,i} + e_i \sim \chi^2(M-1)$$

  - Test statistic: $H = nR^2$
  - Reject if $H > \chi^2_\alpha(M-1)$
- Alternatively, use the Hausman test
  - H_0: X_i is exogenous, i.e. OLS and 2SLS are both consistent
  - Test statistic: $H = \frac{(\hat{\beta}_{1,2SLS} - \hat{\beta}_{1,OLS})^2}{\mathrm{var}(\hat{\beta}_{1,2SLS} - \hat{\beta}_{1,OLS})} \sim \chi^2(1)$
  - Reject if $H > \chi^2_\alpha(1)$
- IV is consistent if instrument is relevant (F-test > 10), but bias can be large

$$\text{Bias IV} \sim \frac{\{\#\ \text{instruments}\ \} \times \rho(U_i, V_i) \times \left(1 - R^2_{\text{partial}}\right)}{\{\#\ \text{observations}\ \} \times R^2_{\text{partial}}}$$

where R^2_{partial} is the contribution of the instruments to R^2 in the first-stage

Lecture 2: IV

- IV is weak if $\mathbf{cov}(Z_i, X_i)$ is small. Recall

$$\mathbf{plim}_{n\to\infty}\hat{\beta}_{1,2SLS} = \frac{\mathbf{cov}(Z_i, Y_i)}{\mathbf{cov}(Z_i, X_i)}$$

- When $\mathbf{cov}(Z_i, X_i)$ is close to 0, i.e. instrument is irrelevant, then the sampling variation in $\mathbf{cov}(Z_i, X_i)$ is not helpful to estimate $\beta_{1,2SLS}$
- Weak instruments can be detected in the first-stage using a t-test or a F-test
    - Rule of thumb: instrument is weak if bias IV is larger than 10% of the bias of OLS

$$\frac{\text{Bias IV}}{\text{Bias OLS}} \approx \frac{\{\# \text{ instruments }\}}{\{\# \text{ observations }\} \times R^2_{\text{partial}}}$$

- Do you want to study more about weak IV? Take the field course Advanced Microeconometrics

Lecture 3: Panel data models

- Assume N individuals observed over T periods

$$Y_{it} = \alpha + X_{it}'\beta + \eta_i + U_{it}$$

  - η_t is an individual specific effect which captures unobserved heterogeneity
- How to estimate this model?
  - Pooled OLS
  - Fixed-effects model
  - Random-effects model
- Assumptions for all three models
  - Strict exogeneity: $E[U_{it}|X_{i1}, \ldots, X_{iT}, \eta_i)] = 0$ allows for only static panel models
  - Weak exogeneity: $E[U_{it}|X_{it}, \eta_i)] = 0$ allows models to be dynamic
- Do you want to study dynamic panel models? Take the field course Applied Microeconometrics

Lecture 3: Panel data models

- Pooled OLS

$$Y_{it} = \alpha + X_{it}'\beta + U_{it}^*$$

  where $U_{it}^* = \eta_i + U_{it}$
- If $E[\eta_i | X_{i1}, \ldots, X_{iT})] \neq 0$, i.e. individual specific effects are correlated with regressors, the OLS estimator of β is biased and inconsistent
- If $E[\eta_i | X_{i1}, \ldots, X_{iT}] = 0$, the OLS estimator of β is unbiased and consistent, but we still have that $E[U_{it}^* U_{is}^*] \neq 0 \implies$ use clustered s.e.

Lecture 3: Panel data models

- Fixed-effects model

$$Y_{it} = \alpha + X_{it}'\beta + \eta_i + U_{it}$$

  where η_i is fixed
- Within estimation
  - Estimation

$$\begin{aligned} Y_{it} - \bar{Y}_i &= \alpha + X_{it}'\beta + \eta_i + U_{it} - (\alpha + \bar{X}_i{}'\beta + \eta_i + \bar{U}_i) \\ &= (X_{it} - \bar{X}_i)'\beta + (U_{it} - \bar{U}_i) \end{aligned}$$

  - Assumption: $E[(X_{it} - \bar{X}_i)'(U_{it} - \bar{U}_i)] = 0$
- First-difference
  - Estimation

$$\begin{aligned} Y_{it} - Y_{it-1} &= \alpha + X_{it}'\beta + \eta_i + U_{it} - (\alpha + X_{it-1}'\beta + \eta_i + U_{it-1}) \\ &= (X_{it} - X_{it-1})'\beta + (U_{it} - U_{it-1}) \end{aligned}$$

  - Assumption: $E[(X_{it} - X_{it-1})'(U_{it} - U_{it-1})] = 0$
- Do you want to combine IV and FE? Take the core course Advanced Econometrics II

Lecture 3: Panel data models

- If strict exogeneity is violated, both FE estimators are not consistent
- To test strict exogeneity, use the following specifications
  - For $T = 2$, both estimators are the same so check

$$Y_{it} - Y_{it-1} = (X_{it} - X_{it-1})'\beta + X_{it}'\gamma + (U_{it} - U_{it-1})$$

  - $H_0 : \gamma = 0$, use a t-test or F-test to check that
  - For $T > 2$, the estimators should be close

$$Y_{it} = \alpha + X_{it}'\beta + X_{it+1}'\gamma + \eta_i + U_{it}$$

  - $H_0 : \gamma = 0$, use a t-test or F-test to check that

Lecture 3: Panel data models

- To check serial correlation, estimate the first difference regression

$$\begin{aligned} Y_{it} - Y_{it-1} &= (X_{it} - X_{it-1})'\beta + (U_{it} - U_{it-1}) \\ &= (X_{it} - X_{it-1})'\beta + E_{it} \\ \implies \hat{E}_{it} &= (Y_{it} - Y_{it-1}) - (X_{it} - X_{it-1})'\hat{\beta} \end{aligned}$$

- Estimate the following model

$$\hat{E}_{it} = \rho \hat{E}_{it-1} + e_{it}$$

- $H_0 : \rho = -0.5$, use a t-test to check that. If there is no autocorrelation

$$\begin{aligned} \hat{\rho} &= \frac{\text{cov}(E_{it-1}, E_{it})}{\text{cov}(E_{it-1}^2)} = \frac{\text{cov}(U_{it-1} - U_{it-2}, U_{it} - U_{it-1})}{\text{cov}(U_{it} - U_{it-1}, U_{it} - U_{it-1})} \\ &= \frac{\text{cov}(U_{it-1}, U_{it}) - \text{cov}(U_{it-1}, U_{it-1}) - \text{cov}(U_{it-2}, U_{it}) + \text{cov}(U_{it-2}, U_{it-1})}{\text{cov}(U_{it}, U_{it}) - \text{cov}(U_{it}, U_{it-1}) - \text{cov}(U_{it-1}, U_{it}) + \text{cov}(U_{it-1}, U_{it-1})} \\ &= \frac{-\text{cov}(U_{it-1}, U_{it-1})}{\text{cov}(U_{it}, U_{it}) + \text{cov}(U_{it-1}, U_{it-1})} = \frac{-\sigma_u^2}{2\sigma_u^2} = -\frac{1}{2} \end{aligned}$$

- If there is autocorrelation, use robust s.e.

Lecture 3: Panel data models

- Random-effects model

$$Y_{it} = \alpha + X_{it}'\beta + \eta_i + U_{it}$$

  where η_i is random and $E(\eta_i | X_{i1}, \ldots, X_{iT}) = 0$
- Estimation
  - Stack observations of all individuals $Y_i = X_i'\beta + e_T\eta_i + U_i$
- If σ_η^2 and σ_u^2 known, use the GLS estimator

$$\hat{\beta}_{GLS} = \sum_{i=1}^{N}(X_i'\Omega^{-1}X_i)^{-1}\sum_{i=1}^{N}(X_i'\Omega^{-1}Y_i)$$

  where $\mathbf{var}(e_T\eta_i + U_i) = \sigma_u^2(I_T + \frac{\sigma_\eta^2}{\sigma_u^2}e_Te_T') = \sigma_u^2\Omega$
- If σ_η^2 and σ_u^2 unknown, use the FGLS estimator
  - Estimate σ_u^2 by within estimation
  - Estimate σ_η^2 by between estimation
  - Do GLS with $\hat{\Omega}$ instead of Ω
- In general σ_η^2 and σ_u^2 are unknown, so one has to apply FGLS

Lecture 3: Panel data models

- FE or RE model
  - RE can deal with time-invariant regressors
  - RE can be used to make predictions outside the sample
  - RE has a stronger assumption that $E(\eta_i | X_{i1}, \ldots, X_{iT}) = 0$
  - FE robust against correlation of individual effects and regressors
  - FE is less efficient and parameter estimates might be noisy
- Use the Mundlak procedure, to test which model to use
  - Estimate the RE model

$$Y_{it} = X_{it}'\beta + \bar{X}_i'\gamma + \omega_i + U_{it}$$

    where $\eta_i = \bar{X}_i\gamma + \omega_i$ and ω_i is a random effect that is uncorrelated with X_{it}
  - $H_0 : \gamma = 0$, i.e. the random effects model should be used
- Alternative use the Hausman test

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\text{var}(\hat{\beta}_{FE}) - \text{var}(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \sim \chi^2(R)$$

  where R is the number of time-varying regressors
  - $H_0 : E(\eta_i | X_{i1}, \ldots, X_{iT}) = 0$, i.e. RE and FE are consistent, but RE is more efficient

Lecture 4: Potential outcomes model

- The goal of policy evaluation is to obtain a causal effect of treatment on the outcome of interest
- Potential outcomes model
  - Each individual has 2 potential outcomes: Y_{1i}^* if treated and Y_{0i}^* if untreated
  - $\Delta_i = Y_{1i}^* - Y_{0i}^*$ - individual effect of participating in treatment (not observed)
  - This is the fundamental problem of causal inference
- Treatment effects

$$ATE = E[\Delta] = E[Y_1^* - Y_0^*] = E[Y_1^*] - E[Y_0^*]$$

  - ATE is the effect for the full population

$$ATET = E[\Delta|D = 1] = E[Y_1^* - Y_0^*|D = 1] = E[Y_1^*|D = 1] - E[Y_0^*|D = 1]$$

  - ATET is the effect for individuals who actually received the treatment

Lecture 4: Potential outcomes model

- If there is self-selection into treatment, participation might not be independent of the potential outcomes, i.e. people with positive individual effects are more likely to participate

$$E[Y_1^*] \neq E[Y_1^*|D=1]$$
$$E[Y_0^*] \neq E[Y_0^*|D=0]$$

- In this case
  - $E[Y_1^*|D=1]$ and $E[Y_0^*|D=0]$ can be estimated
  - $E[Y_1^*|D=0]$ and $E[Y_0^*|D=1]$ can't be estimated
- A solution is to use randomized experiments
  - Treatment is assigned randomly, i.e $(Y_{0i}^*, Y_{1i}^*) \perp D_i$
  - So we can assume the same expected effect for treated and untreated

$$E[Y_1^*] = E[Y_1^*|D=1] = E[Y_1^*|D=0]$$
$$E[Y_0^*] = E[Y_0^*|D=0] = E[Y_0^*|D=1]$$

which implies $ATE = ATET$

Lecture 4: Potential outcomes model

- To estimate the treatment effect, let the observed outcome be

$$Y_i = D_i Y_{1i}^* + (1 - D_i) Y_{0i}^*$$

where $D_i = 1$ if a person received treatment

  - Estimate the sample means to get the estimators

$$E(\hat{Y_1^*}|D = 1) = \frac{\sum_{i=1}^{N} D_i Y_i}{\sum_{i=1}^{N} D_i}$$

$$E(\hat{Y_0^*}|D = 0) = \frac{\sum_{i=1}^{N}(1 - D_i) Y_i}{\sum_{i=1}^{N}(1 - D_i)}$$

$$A\hat{T}E = AT\hat{E}T = \frac{\sum_{i=1}^{N} D_i Y_i}{\sum_{i=1}^{N} D_i} - \frac{\sum_{i=1}^{N}(1 - D_i) Y_i}{\sum_{i=1}^{N}(1 - D_i)}$$

- Potential outcomes model is equivalent to the difference-in-means estimator

$$Y_i = \alpha + \delta D_i + U_i$$

where $\delta = A\hat{T}E = AT\hat{E}T$

Lecture 4: Potential outcomes model

- Validity of experiments
  - Internal validity (no spill-over effects, no substitution, no Hawthorne effect) - extent to which we can make causal inference
  - External validity - how experimental results generalize
  - Stable unit treatment value assumption (SUTVA) - treatment participation of one individual does not affect the potential outcomes of other individuals
- Field experiments used because randomized experiments are rare - usually implemented as randomization in natural environment
- Types of field experiments
  - Oversubscription - if there are more applicants than available slots, assign treatment by lottery
  - Phasing-in - start low scale, expand later
  - Within-group - only some subgroups get treatment, others not
  - Encouragement design - randomly encourage subsample to participate
- DiNardo and Lee (2011) judge every method for evaluation on three criteria
  - Appropriate description of treatment assignment mechanism
  - Consistent with wide class of behavioral models
  - Yields testable implications

Lecture 5: LATE and power analysis

- If there is partial compliance in an experiment, initial treatment assignment is often not equal to actual treatment assignment, i.e. $Z_i \neq D_i$
- If people self-select into treatment, treatment effect is heterogeneous

$$Y_i = \alpha + \delta_i D_i + U_i$$

  where δ_i is individual (heterogeneous) effect
- In this case Imbens and Angrist (1994) suggest to study LATE imposing monotonicity assumption
  - If you get the treatment, you don't want to opt out
  - If you don't get the treatment, you want to get one

$$D_i(1) \geq D_i(0)$$

- When the instrument is binary, LATE is equal to

$$LATE = \frac{E[Y|Z=1] - E[Y|Z=0]}{Pr[D=1|Z=1] - Pr[D=1|Z=0]}$$

Lecture 5: LATE and power analysis

- If monotonicity holds, LATE is the average treatment effect for compliers
  - Without monotonicity - difficult interpretation
- Randomization implies the same share of compliers, always takers, and never takers in treated and control groups
  - Compliers: $D(1) = 1$ and $D(0) = 0$
  - Always takers: $D(1) = D(0) = 1$
  - Never takers: $D(1) = D(0) = 0$
  - Defiers: $D(1) = 0$ and $D(0) = 1$ - ruled out by monotonicity

Lecture 5: LATE and power analysis

- What is the minimum effect we are able to detect, given the treatment intensity, the number of participants, and the power?

$$MDE = (t_{1-\alpha/2} - t_{1-q})\sqrt{\frac{1}{p(1-p)}\frac{\sigma^2}{n}}\frac{1}{r_t - r_c}$$

- What is smallest sample size needed to run an experiment, given MDE, the treatment intensity, and the power?

$$n = \left(\frac{t_{1-\alpha/2} - t_{1-q}}{MDE}\right)^2 \frac{\sigma^2}{p(1-p)} \left(\frac{1}{r_t - r_c}\right)^2$$

  where MDE - minimum detectable effect; n - sample size; p - treatment intensity; σ^2 - variance; r_t - compliance rate in the treatment group; r_c - treatment intensity in the control group
- MDE can be based on
  - earlier literature
  - requirements from partner
  - cost-benefit analysis

Lecture 6: DiD

- To compare differences between treatment and control groups before and after the intervention, estimate the following model

$$Y_{gt} = \alpha_t + \eta_g + \delta D_{gt} + U_{gt}$$

  where α_t - common time trend, η_g - group specific effect
- If we have 2 groups and 2 time periods, we can rewrite that as

$$\begin{aligned} Y_{T0} &= \alpha_0 + \eta_T + U_{T0} \\ Y_{T1} &= \alpha_1 + \delta + \eta_T + U_{T1} \\ Y_{C0} &= \alpha_0 + \eta_C + U_{C0} \\ Y_{C1} &= \alpha_1 + \eta_C + U_{C1} \end{aligned}$$

  Take the differences of expected values

$$\begin{aligned} E[Y_{T1}] - E[Y_{T0}] &= (\alpha_1 + \delta + \eta_T) - (\alpha_0 + \eta_T) = \alpha_1 + \delta - \alpha_0 \\ E[Y_{C1}] - E[Y_{C0}] &= (\alpha_1 + \eta_C) - (\alpha_0 + \eta_C) = \alpha_1 - \alpha_0 \\ \text{DiD} &= (\alpha_1 + \delta - \alpha_0) - (\alpha_1 - \alpha_0) = \delta \end{aligned}$$

- DiD estimates ATET if there is a constant treatment effect δ or there are only 2 periods

Lecture 6: DiD

- Key assumption: parallel trend assumption
- Parallel trend assumption is scale dependent
  - If prior trends are the same in the logarithm of wages, they are not equal in wage levels
- Intervention should be random conditional on time and group specific effects, otherwise $E[U_{g0}|D_g] \neq 0$ which violates exogeneity
  - Example: Ashenfelter dip - treatment participants have a dip in outcomes just before entering the programme
- How to test the parallel trend assumption?
  - Check prior trends
  - Do placebo checks
- If there is no support for parallel trend assumpition
  - Include time-varying covariates or group specific trends
  - DDD
  - Synthetic control group
  - DiD with IV
  - Changes-in-Changes

Lecture 6: DiD

- If there are more than two periods, we can implement an event-study specification

$$Y_{gt} = \alpha_t + \eta_g + \delta_{t-\tau_g} D_{gt} + U_{gt}$$

  where
  - τ_g is the moment of the treatment
  - t is time period
- Impose the normalisation $\delta_{-1} = 0$, which is the coefficient for the last period before the treatment, otherwise you get perfect multicollinearity
- If data sampling process or treatment is clustered, use clustered s.e.
  - Abadie et al. 2017 discuss clustering
- Do you want to discuss clustering s.e. more? Take the core course Advanced Econometrics II

Lecture 7: RDD and RKD

- Sharp RDD
  - Treatment assignment is sharp at the cutoff point $\bar{S}$

$$D_i = I(S_i > \bar{S})$$

  - Use crossing the cutoff, to estimate the marginal treatment effect

$$MTE(\bar{S}) = \lim_{s \downarrow \bar{S}} E[Y_i | S_i = s] - \lim_{s \uparrow \bar{S}} E[Y_i | S_i = s]$$

- The model for sharp RDD to estimate the effect of a treatment on the outcome

$$Y_i = \alpha + \delta D_i + K(S_i - \bar{S}) + U_i$$

- RDD and RKD exploit local randomization
  - Interpretation: ATE for people who change status when moving from just below to just above $\bar{S}$ - LATE

Lecture 7: RDD and RKD

- Fuzzy RDD
  - Treatment assignment is discontinuous at the cutoff point $\bar{S}$

$$\lim_{s \downarrow \bar{S}} P(D_i = 1 | S_i = s) \neq \lim_{s \uparrow \bar{S}} P(D_i = 1 | S_i = s)$$

  - Use crossing the cutoff as a locally valid instrument, to estimate the marginal treatment effect

$$MTE(\bar{S}) = \frac{\lim_{s \downarrow \bar{S}} E[Y_i | S_i = s] - \lim_{s \uparrow \bar{S}} E[Y_i | S_i = s]}{\lim_{s \downarrow \bar{S}} P[D_i = 1 | S_i = s] - \lim_{s \uparrow \bar{S}} P[D_i = 1 | S_i = s]}$$

- The model for fuzzy RDD uses two stages as in IV
  - First-stage: estimate the effect of crossing $\bar{S}$ on the probability to get the treatment

$$D_i = \gamma_0 + \gamma_1 I(S_i > \bar{S}) + G(S_i - \bar{S}) + V_i$$
$$\implies \hat{D}_i = \hat{\gamma}_0 + \hat{\gamma}_1 I(S_i > \bar{S}) + \hat{G}(S_i - \bar{S})$$

  - Second-stage: use $\hat{D}_i$ to estimate the effect of the treatment on the outcome

$$Y_i = \alpha + \delta \hat{D}_i + K(S_i - \bar{S}) + U_i$$

Lecture 7: RDD and RKD

- RKD is similar to RDD, but instead of a jump in the intercept, the slope changes at the threshold
- The model for RKD uses "two stages"
  - First-stage: estimate the effect of crossing $\bar{S}$ on the probability to get the treatment

$$D_i = \gamma_0 + \gamma_1(S_i - \bar{S})I(S_i < \bar{S}) + \gamma_2(S_i - \bar{S})I(S_i \geq \bar{S}) + V_i$$

  - Second-stage: estimate the effect of crossing $\bar{S}$ on the changes in the slope

$$Y_i = \beta_0 + \delta_1(S_i - \bar{S})I(S_i < \bar{S}) + \delta_2(S_i - \bar{S})I(S_i \geq \bar{S}) + U_i$$

  - Estimate the causal effect of D_i on Y_i at $S_i = \bar{S}$

$$\frac{\delta_2 - \delta_1}{\gamma_2 - \gamma_1}$$

- There is a treatment effect if

$$\lim_{S_i \uparrow \bar{S}} \frac{\partial E(Y_i|S_i)}{\partial S_i} = \delta_1 \neq \delta_2 = \lim_{S_i \downarrow \bar{S}} \frac{\partial E(Y_i|S_i)}{\partial S_i}$$

Lecture 7: RDD and RKD

- Good practices
    - Use the McCrary test to check for continuity of density of S_i around the cutoff $\bar{S}$
    - Choose different bandwidths to check sensitivity
    - Choose different functional forms but don't use higher-order polynomials
    - Check if other characteristics are balanced around the discontinuity
    - Use controls as placebo tests
    - Try to use local-linear regression instead of polynomials
- Threats to validity
    - Treatment assignment rule may be public knowledge, which may trigger behavioral responses
    - Possible manipulation of the treatment variable

Courses

- Advanced Econometrics I
- Advanced Econometrics II
- Advanced Microeconometrics
- Applied Microeconometrics

Final thoughts

- Good luck :)