Summary of Introduction to Econometrics
Prof. Alam

1 Population concepts

For random variables X and Y and constants a and b, and functions $f()$ and $g()$:

1.1 Expectation

Sample counterpart of $E(X)$ is the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

- $E(aX + bY) = aE(X) + bE(Y)$
- $E(X + Y) = E(X) + E(Y)$
- $E(a) = a$
- $E(XY) = E(X)E(Y)$ if X and Y are independent
- $E(XY) = E(X)E(Y) + \text{Cov}(X, Y)$
- $E(X|Y) = E(X)$ if X and Y are independent
- $E(g(Y) + f(X)) = E(g(Y)) + E(f(X))$
- $E(g(Y) + f(X) \mid X) = E(g(Y) \mid X) + f(X)$

1.2 Covariance

Sample counterpart of $\text{Cov}(X, Y)$ is the sample covariance $s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

- $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = 0$ if X and Y are independent
- $\text{Cov}(f(X), g(Y) \mid X) = 0$
- $\text{Cov}(X, X) = E(X^2) - E(X)^2 = \text{Var}(X)$

1.3 Variance

Sample counterpart of $\text{Var}(X)$ is the sample variance $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

- $\text{Var}(X) = E(X^2) - E(X)^2$
- $\text{Var}(a) = 0$
- $\text{Var}(aX) = a^2\text{Var}(X)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$
- $\text{Var}(g(Y) + f(X)) = \text{Var}(g(Y)) + \text{Var}(f(X)) + 2\text{Cov}(g(Y), f(X))$
- $\text{Var}(g(Y) + f(X) \mid X) = \text{Var}(g(Y) \mid X)$

Sample Identity	Population Counterpart
$\sum(x_i - \bar{x}) = 0$	$E(X - E(X)) = 0$
$\sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$	$E((X - E(X))^2) = E(X^2) - (E(X))^2$
$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i(y_i - \bar{y}) = \sum y_i(x_i - \bar{x})$	$E((X - E(X))(Y - E(Y))) = E(X(Y - E(Y))) = E(Y(X - E(X)))$

Table 1: Sample identities and their population counterparts

- Law of Iterated Expectations: $E(X) = E(E(X|Y))$
- Law of Large Numbers: As $n \to \infty$, $\bar{x} \to E(X)$
- Variance Decomposition Formula: $\mathrm{Var}(X) = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y))$

2 Simple Linear Regressions

Population model: $y_i = \beta_0 + \beta_1 x_i + u_i$

2.1 Assumptions SLR 1- 5

- SLR 1: Linearity: In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as: $y = \beta_0 + \beta_1 x + u$ where β_0 and β_1 are the population intercept and slope parameters, respectively.
- SLR 2: Random sampling: We have a random sample of size n,$\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ from the population model
- SLR 3: Sample variation in X: The sample explanatory x variable, namely, x_i , $i \in \{1, \cdots, n\}$, are not all the same value $\widehat{Var}(x_i) > 0$
- SLR 4: Zero conditional mean: The error u has an expected value of zero given any value of the explanatory variable. In other words, $E(u_i \mid x_i) = 0$
- SLR 5: Homoskedasticity: The error u has the same variance given any value of the explanatory variable. In other words, $Var(u_i \mid x_i) = \sigma^2$

2.2 OLS Estimates

The OLS estimates are $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squared residuals, $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$, $s_{xy} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, and $s_{xx} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$.
In class we showed that, we can also write $\hat{\beta}_1$ as:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

2.3 Theorems

2.3.1 Unbiasedness of the OLS estimator

Under Assumptions SLR. 1 through SLR.4, $\hat{\beta}_0$ is unbiased for β_0, and $\hat{\beta}_1$ is unbiased for β_1. That is,

$$\mathrm{E}\left(\hat{\beta}_0\right) = \beta_0 \text{ and } \mathrm{E}\left(\hat{\beta}_1\right) = \beta_1$$

2.3.2 Variance of the OLS estimator

Denote σ^2 as the $Var(u_i)$. Under Assumptions SLR. 1 through SLR.5,

$$\operatorname{Var}\left(\hat{\beta}_1\right)=\frac{\sigma^2}{\sum_{i=1}^n\left(x_i-\bar{x}\right)^2}=\frac{\sigma^2}{\mathrm{SST}_x}, \quad \text{and} \quad \operatorname{Var}\left(\hat{\beta}_0\right)=\frac{\sigma^2 \frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n\left(x_i-\bar{x}\right)^2}=\frac{\sigma^2 \frac{1}{n} \sum_{i=1}^n x_i^2}{\mathrm{SST}_x}$$

where these are conditional on the sample values $\{x_1, \ldots, x_n\}$ and $SST_x=\sum_{i=1}^n\left(x_i-\bar{x}\right)^2$.

2.3.3 Unbiased estimate of σ^2

Under Assumptions SLR.1 through SLR.5, $\mathrm{E}\left(\hat{\sigma}^2\right)=\sigma^2$, where $\hat{\sigma}^2=\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2$, and $\hat{u}_i=y_i-\hat{\beta}_0-\hat{\beta}_1 x_i$

2.4 Standard errors of the OLS estimates

The standard error of the OLS estimate on the slope $\hat{\beta}_1$ is given by:

$$se(\hat{\beta}_1)=\sqrt{Var(\hat{\beta}_1)}=\sqrt{\frac{\hat{\sigma}^2}{\mathrm{SST}_x}}$$

3 Multiple Linear Regressions

Population model: $y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_k x_{ki}+u_i$

3.1 Assumptions MLR 1-6

- MLR 1: Linearity: The population model is linear in the parameters: $y=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k+u$
- MLR 2: Random sampling: We have a random sample of size n,$\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i): i=1,2, \ldots, n\}$ from the population model
- MLR 3: No perfect multicollinearity: The explanatory variables are not perfectly collinear, i.e., there are no exact linear relationships among the explanatory variables.
- MLR 4: Zero conditional mean: The error u has an expected value of zero given any values of the explanatory variables. In other words, $E(u_i \mid x_{1i}, x_{2i}, \ldots, x_{ki})=0$
- MLR 5: Homoskedasticity: The error u has the same variance given any values of the explanatory variables. In other words, $Var(u_i \mid x_{1i}, x_{2i}, \ldots, x_{ki})=\sigma^2$
- MLR 6: Normality: The error term u_i is normally distributed given any values of the explanatory variables.

3.2 OLS Estimates

The OLS estimates are $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ that minimize the sum of squared residuals, $\sum_{i=1}^n \hat{u}_i^2=\sum_{i=1}^n\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{1i}-\ldots-\hat{\beta}_k x_{ki}\right)^2$

3.3 Theorems

3.3.1 Unbiasedness of the OLS estimator

Under Assumptions MLR. 1 - MLR.4, $\hat{\beta}_0$ is unbiased for β_0, and $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are unbiased for $\beta_1, \ldots, \beta_k$. That is,

$$\mathrm{E}\left(\hat{\beta}_0\right)=\beta_0, \quad \mathrm{E}\left(\hat{\beta}_1\right)=\beta_1, \ldots, \mathrm{E}\left(\hat{\beta}_k\right)=\beta_k$$

3.3.2 Variance of the OLS estimator

Under assumptions MLR.1 - MLR.5, the sampling variance of the OLS estimate on the j^{th} coefficient $\hat{\beta}_j$ is given by:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{(1 - R_j^2)\underbrace{\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2}_{SST_{x_j}}} = \frac{\sigma^2}{(1 - R_j^2)SST_{x_j}}$$

for $j = 0, 1, \ldots, k$. where:

- these are conditional on the sample values $\{x_{1i}, x_{2i}, \ldots, x_{ki}\}$
- σ^2 is the variance of the error term u_i conditional on the independent variables
- $\bar{x}_j$ is the sample mean of x_{ji}
- $\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2$ is the total sample variance of x_{ji}
- R_j^2 is the R^2 from regressing x_{ji} on all other independent variables and an intercept.

3.3.3 Unbiased estimate of σ^2

Under Assumptions MLR.1 - MLR.5, E $\left(\hat{\sigma}^2\right) = \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2$, and $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \ldots - \hat{\beta}_k x_{ki}$

3.4 Standard errors of the OLS estimates

The standard error of the OLS estimate on the j^{th} coefficient $\hat{\beta}_j$ is given by:

$$se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} = \sqrt{\frac{\hat{\sigma}^2}{(1 - R_j^2)SST_{x_j}}}$$

3.5 Ommited Variable Bias

- For the true population model satisfying MLR1-MLR4:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

- Suppose we omitted x_{2i} and only ran the simple regression of y_i on x_{1i}:

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1i}$$

- $\tilde{\beta}_1$ relates to $\hat{\beta}_1$ via:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \delta_1$$

  where δ_1 is the slope from the regression of x_{2i} on x_{1i} from estimating:

$$x_{2i} = \delta_0 + \delta_1 x_{1i} + \epsilon_i$$

- So $E(\tilde{\beta}_1) = \beta_1 + \beta_2 \delta_1$.
- Bias = $E(\tilde{\beta}_1) - \beta_1 = \beta_2 \delta_1$

4 Inference

Central Limit Theorem (CLT): Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of size n from a population with mean μ and variance σ^2. Then, as n approaches infinity, the distribution of sample means is a normal distribution. i.e., $\bar{X} \sim N(\mu, \sigma^2/n)$ so that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

4.1 Single coefficient hypothesis test

Under the null hypothesis $H_0 : \beta_j = b$, the test statistic is

$$t = \frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)}$$

- The $(1-\alpha)$ CI for β_j: $\hat{\beta}_j \pm t^*_{\alpha/2} se(\hat{\beta}_j)$ where $t^*_{\alpha/2}$ is the critical value of the t-distribution at the $\alpha/2$ level.
- As sample size becomes large t distribution converges to standard normal distribution.
- Table for critical t-values in large samples for two-tailed tests at various significance levels α:

Significance level	0.10	0.05	0.025	0.01	0.005	0.001
Critical value	1.645	1.96	2.576	2.807	3.291	3.922

4.2 Multiple coefficient hypothesis test

Testing q restrictions, under the null hypothesis $H_0 : \beta_{j_1} = \ldots = \beta_{j_q} = 0$, the test statistic is

$$F = \frac{SSR_{UR} - SSR_R}{q} \Bigg/ \frac{SSR_R}{n-k-1} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k-1)}$$

where subscripts UR and R represent the unrestricted and restricted model respectively, $n-k-1$ are the degrees of freedom of the UR model, and SSR represents the sum of squared residuals.

4.3 Rejection rules at a given α level

- If test statistic is higher* than the critical value (careful on one sided tests)
- If p-value is lower than the significance level
- If the confidence interval does not contain the null hypothesis value

5 Variants on MLR

Not much formulae to give here. Study slides.

6 Potential Outcomes Framework

- Treatment D_i for each unit i with observed outcome Y_i
- We observe only *one of the two* **potential outcomes** $Y_i(0)$ or $Y_i(1)$

$$\underbrace{E\left[\mathrm{Y}_i \mid \mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_i \mid \mathrm{D}_i = 0\right]}_{\text{Observed difference in avg outcomes}} = \quad E\left[\mathrm{Y}_i(1) \mid \mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_i(0) \mid \mathrm{D}_i = 0\right]$$

$$= \underbrace{E\left[\mathrm{Y}_i(1) \mid \mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_i(0) \mid \mathrm{D}_i = 1\right]}_{\text{average treatment effect on the treated}} \underbrace{+ E\left[\mathrm{Y}_i(0) \mid \mathrm{D}_i = 1\right] - E\left[\mathrm{Y}_i(0) \mid \mathrm{D}_i = 0\right]}_{\text{selection bias}}$$

7 Randomized Control Trials

- Random assignment of $D_i \implies E[Y_i(0) \mid D_i = 1, X_i] = E[Y_i(0) \mid D_i = 0, X_i]$ because $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \mid X_i$
- $E(\varepsilon_i | D_i) = 0 \implies cov(D_i, \varepsilon_i) = 0$
- Bunch of things to worry about and take care of while designing a RCT:
  - Sample size
  - Stratified randomization: Balance of covariates
  - Spillover effects

8 Instrumental Variables, 2SLS and heterogeneous treatment effects

Failure of MLR 4: $E(\varepsilon_i|D_i) \neq 0 \implies cov(D_i, \varepsilon_i) \neq 0$: D_i **is endogenous.**
$\widehat{\beta_1}^{OLS} = \frac{\widehat{cov}(Y_i,D_i)}{\widehat{var}(D_i)}$ will be **biased** & **inconsistent** when D_i is **endogenous**

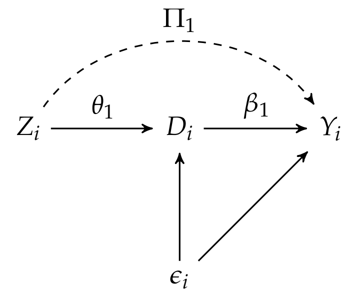
Valid IV Z_i satisfies:

- $cov(Z_i, D_i) \neq 0$ (IV relevance) : testable
- $cov(Z_i, \varepsilon_i) = 0$ (IV exogeneity) : non-testable

Structural equation: $Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$

First stage: $D_i = \theta_0 + \theta_1 Z_i + u_i$

Reduced form: $Y_i = \Pi_0 + \Pi_1 Z_i + e_i$



$\widehat{\beta}_{IV} = \frac{\widehat{\Pi}_1}{\widehat{\theta}_1} = \frac{\widehat{cov}(Y_i,Z_i)}{\widehat{cov}(Z_i,D_i)}$
With binary treatment and outcome this boils down to the Wald estimator:

$$\widehat{\beta}_{IV} = \frac{\widehat{\Pi}_1}{\widehat{\theta}_1} = \frac{\widehat{E}(Y_i|Z_i=1) - \widehat{E}(Y_i|Z_i=0)}{\widehat{Pr}(D_i=1|Z_i=1) - \widehat{Pr}(D_i=1|Z_i=0)}$$

Weak IV:
Asymptotics, as $N \to \infty$,

$$\widehat{\beta}_{IV} \to \beta_1 + \frac{cov(z_i, \epsilon_i)}{cov(z_i, D_i)}$$

$$\widehat{\beta}_{OLS} \to \beta_1 + \frac{cov(D_i, \epsilon_i)}{cov(D_i, D_i)}$$

The ratio of the asymptotic bias of IV to that of OLS equals

$$\frac{cov(z_i, \epsilon_i) cov(z_i, D_i)}{cov(D_i, \epsilon_i) cov(D_i, D_i)} = \frac{\rho_{z\epsilon}}{\rho_{D\epsilon}\rho_{zD}}$$

2SLS estimator:

- Structural equation:

$$y = \beta_0 + \beta_1 x_{1,i} + ... + \beta_K x_{K,i} + \beta_D \underbrace{D_i}_{endogenous\ variable} + \varepsilon_i$$

- First stage: regress the endogenous variable on all **the exogenous covariates**

$$D_i = \delta_0 + \delta_1 x_{1,i} + ... + \delta_K x_{K,i} + \theta Z_i + u_i$$

- Generate predicted values of the endogenous variable:

$$\widehat{D}_i = \widehat{\delta}_0 + \widehat{\delta}_1 x_{1,i} + ... + \widehat{\delta}_K x_{K,i} + \widehat{\theta} Z_i$$

- The variance of the 2SLS estimator for the coefficient on D_i is:

$$\text{Var}(\hat{\beta}_{2SLS}) = \frac{\hat{\sigma}^2/(n-K)}{R^2_{D,Z} \cdot (1 - R^2_{D,X}) \cdot \sum_{i=1}^{n}(D_i - \bar{D})^2}$$

  - $R^2_{D,Z}$ is the partial R^2 from the first stage after controlling for exogenous covariates X
  - $R^2_{D,X}$ is the R^2 from regressing D on exogenous covariates X
  - $\hat{\sigma}^2$ is the estimated residual variance from the second stage regression

Multiple IVs (and multiple endogenous variables):

- Each IV must separately satisfy relevance and exogeneity
- No IV can be a linear combination of the others
- The number of IVs must be at least as many as the number of endogenous variables

Heterogeneous treatment effects:

- We can express both the numerator and denominator of IV estimator using the four groups and their potential outcomes $\{Y_i(0), Y_i(1)\}$

Groups	$Z = 0$	$Z = 1$
Never takers (NT)	$D = 0$	$D = 0$
Defiers (DF)	$D = 1$	$D = 0$
Compliers (C)	$D = 0$	$D = 1$
Always takers (AT)	$D = 1$	$D = 1$

- IV identifies and estimates the Local Average Treatment Effect (LATE) which is the ATE for the compliers (C):

$$\text{LATE} = E[Y_i(1) - Y_i(0) \mid C]$$

- In addition to exogeneity:
  - The first assumption needed is of monotonicity (capturing relevance):
    * The IV can only increase or only decrease the prob. of participation for all units
    - $\Pr(D_i = 1 \mid Z_i = 1) > \Pr(D_i = 1 \mid Z_i = 0)$
    - It means that $Pr(DF) = 0$.
  - The second assumption is that there are some compliers, $Pr(C) > 0$