ECONOMETRICS

HONOR'S EXAM REVIEW SESSION

Eunice Han

ehan@fas.harvard.edu

March 26th, 2013

Harvard University

Information

- Exam: April 3rd 3-6pm @ Emerson 105
- Bring a calculator and extra pens.
- Notes are not allowed.
- There is no surprise in the exam. Use previous exams and problem sets as a guideline.
- The review session only covers the main topics.
- You are responsible for ALL the materials you've learned in class.
- Time management is critical. Do not spend too much time on a single question.
- OH:
  1. March 29th 6-8pm @ Littauer 219
  2. April 1st 1-3pm @ Littauer 212

Topic

1. OLS
   - The Assumptions
   - Omitted Variable Bias
   - Conditional Mean Independence
   - Hypothesis Testing and Confidence Intervals
   - Homoskedasticity vs. Heteroskedasticity
   - Nonlinear Regression Models: Polynomials, Log Transformation, and Interaction Terms
2. Binary Dependent Variables
   - Linear Probability Model
   - Probit Model and Logit Model
   - Ordered Probit Model
3. Panel Data
   - Fixed Effects: Entity FE and Time FE
   - Serial Correlation and Clustered HAC SE

Topic (Cont.)

4. Internal Validity and External Validity
   - Threats to internal validity
   - Threats to internal validity
5. Instrumental Variables Regression
   - Conditions for Valid Instruments: Relevance and Exogeneity
   - Tests of Instrumental Validity: F-test and J-test
   - 2SLS estimation: First and Second Stage Regression
6. Time Series Data
   - Stationarity
   - Forecasting Models: AR and ADL Model
   - Dynamic Causal Effects: Distributed Lag Model
   - Serial Correlation and Newey-West HAC SE

Least Squares Assumptions

- Regression Model: $Y_i = \beta_0 + \beta_1 X_i + u_i$

1. (X_i, Y_i) are iid; The sample is drawn randomly from a single population.
2. Large outliers are rare; If this condition fails, OLS estimator is not consistent.
3. $E(u_i | X_i) = 0$; Conditional Mean Zero assumption. Xs are exogenous. This assumption fails if X and u are correlated.
4. No Perfect Multicollinearity (In multivariate regression)

- If all the assumptions are satisfied, the OLS estimates are unbiased and consistent.

Perfect Multicolinearity

- The regressors are said to be perfectly multicollinear if one of the regressors is a perfect linear function of the other regressor(s).
- Suppose $X_{1i} = 2X_{2i}$. Can we estimate regression $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$?
- Using $X_{1i} = 2X_{2i}$, $Y_i = \beta_0 + (2\beta_1 + \beta_2)X_{2i} + u_i$. If we denote $\gamma = 2\beta_1 + \beta_2$, we have $Y_i = \beta_0 + \gamma X_{2i} + u_i$. Here we can estimate unbiased γ using OLS, but we cannot separately estimate both β_1 and β_2.
- Dummy variable trap
  - Ex. If we include both "male" and "female" dummy, perfect multicolinearity occurs because "male"=1- "female".

Univariate Regression Model

- Do married men make more money than single men?
- $lwage_i = \beta_0 + \beta_1 married_i + u_i$
  - lwage= log(monthly wage)
  - married =1 if married, 0 if single

Q1. How would you interpret β_0 and β_1 ?

- β_0(intercept) is the average lwage of single men.
- β_1(slope) represents the difference in average lwage between single and married men.

Q2. In our model, which factors are likely to be in the error terms?

- The error term u_i captures all other factors except marital status that might affect lwage. Education, experience, and ability are potential factors which are omitted here.

Omitted Variable Bias

- Population regression equation (True world)

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Suppose we omitted X_{1i} and estimated the following regression.

$$Y_i = \hat{\beta}_2 X_{2i} + \hat{\varepsilon}_i$$

$$E(\hat{\beta}_2) = \beta_2 + \delta\beta_1 \quad \text{where} \quad \delta = \frac{Corr(X_{1i}, X_{2i})}{Var(X_{2i})}$$

- Now, the OLS estimator is no longer unbiased, and OVB= $\delta\beta_1$

Q1. Under what condition, the OLS estimator suffers from OVB?

1. The omitted variable X_1 is a determinant of Y ($\beta_1 \neq 0$) AND
2. X_1 is correlated with the regressor X_2 ($\delta \neq 0$)

- Can you predict the sign of this OVB?
- What can you do to solve OVB problem?

Multivariate Regression Model

- $lwage_i = \beta_0 + \beta_1 married_i + \beta_2 edu_i + \beta_3 \exp_i + \beta_4 job_i + u_i$

Q1. How would you interpret β_0 and β_1 ?

- β_0 : Predicted value of lwage when all the regressors are zero. Meaningless in this case.
- β_1 : Effect of unit change in "married" on lwage, <u>holding other regressors constant</u>.

Q2. Suppose that we are interested in the effect of marriage on lwage. Can we be sure that the there is no OVB after we include more variables in the regression?

Q3. Suppose we are only interested in the effect of marriage on lwage. Do we really need the strong assumption of exogeneity of Xs, $E(u_i | X_i) = 0$, to obtain unbiased estimator for β_1? NO!

Conditional Mean Independence

- $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \ldots + \beta_{1+r} W_{ri} + u_i$
- X: treatment variable W: control variables.
- If we are only interested in the causal effect of X on Y, we can use a weaker assumption of Conditional Mean Independence: $E(u \mid X, W) = E(u \mid W)$
- The conditional expectation of u does not depend on X if we control for W. Conditional on W, X is as if randomly assigned, so X becomes uncorrelated with u. (but W can still be correlated with u)
- Under the conditional mean independence assumption, OLS can give us the unbiased and consistent estimator for β_1 , but the coefficients for W can be still biased.

Measure of Fit

Q1. What is R^2 and what does it tell us?

- R^2 is the fraction of variation of dependent variable which is explained by our model.

Q2. Why is Adjusted R^2?

- Unlike R^2, adjusted R^2 adjusts for the number of regressors in a model because it increases only if the new regressor improves the model more than would be expected by chance.

Q3. What is SER (Standard Error of Regression)?

- SER estimates the standard deviation of the error term. A large SER implies that the spread of the observations around the regression is large, and there could be other important factors that we have not included in the regression.

Hypothesis Testing

- $H_0 : \beta_1 = 0$ & $H_1 : \beta_1 \neq 0$ (two-sided)
- T-test: By the Central Limit Theorem, t-statistics is normally distributed when n is large enough (Usually N>30).
- t-statistics = $\dfrac{\hat{\beta}_1 - 0}{\sqrt{Var(\hat{\beta}_1)}} \sim N(0,1)$
- If $|t| > 1.96$, we reject null hypothesis at 5% significance level.
- P-value: The p-value is the probability of drawing a value of that differs from 0, by at least as much as the value actually calculated with the data, if the null is true. If p-value is less than 0.05, we reject the null hypothesis at 5% significance level.
- P-value ≈ $\Pr(|z| > |t|) = 2\Phi(-|t|)$

Example for Hypothesis Testing

Q1. Let μ_m be the mean log hourly wage in the population of married men and μ_s be the population mean for single men. Construct null hypothesis and two-sided alternative hypothesis to test the difference between two means. $H_0 : \mu_s - \mu_m = 0$ and $H_1 : \mu_s - \mu_m \neq 0$

Q2. What is the sample mean, the estimator of $\mu_s - \mu_m$? $\bar{Y}_s - \bar{Y}_m$

Q3. Compute SE of the sample mean. $SE(\bar{Y}_s - \bar{Y}_m) = \sqrt{\frac{{S_s}^2}{n_s} + \frac{{S_m}^2}{n_m}}$

Q4. Calculate t-statistics. $t = \frac{(\bar{Y}_s - \bar{Y}_m) - 0}{SE(\bar{Y}_s - \bar{Y}_m)}$

Q5. Suppose that the t-statistic >1.96. What do we learn from this?

- We reject the null at the 5% significance level. Therefore, we learned that married men and single men do not make the same amount of money.

Confidence Interval

- Confidence Interval: An interval that contains the true population parameter with a certain pre-specified probability (usually 95%).
- Confidence Interval is useful because it is impossible to learn the exact value of population mean using the information in one sample due to the random sampling error. However, it is possible to use this data to construct a set of values that contains the true population mean with certain probability.
- How do we construct the 95% confidence interval?
  - $\Pr(-1.96 \leq t \leq 1.96) = 0.95$
  - $[\hat{\beta}_1 - 1.96\ SE(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96\ SE(\hat{\beta}_1)]$
- If confidence interval contains 0, we fail to reject $H_0 : \beta_1 = 0$

Homoskedasticity *vs.* Heteroskedasticity

- Homoskedasticity: The error term u_i is homoskedastic if the variance of the conditional distribution of u_i given X_i is constant for i=1,2,...n.
- Heteroskedasticity: The variance of the conditional distribution of u_i given X_i is different across X_i.

Q. What happens if we have heteroskedasticity problem?

  - The OLS estimator (β_1) is still unbiased and consistent, as long as the OLS assumptions are met (esp. $E(u_i | X_i) = 0$). However, our SE calculated using homoskedasticity-only formula gives us a wrong answer, so the hypothesis testing and confidence intervals based on homoskedasticity-only formula are no longer valid. We have to use <u>heteroskedasticity-robust SE</u>.

Joint Hypothesis Testing

- For joint hypothesis testing, we use F-test.
- Under the null hypothesis, in large samples, the F-statistic has a sampling distribution of $F_{q,\infty}$. That is,
- F-statistic ~ $F_{q,\infty}$ where q is the number of coefficients that you are testing.
- If F-statistics is bigger than the critical value or p-value is smaller than 0.05, we reject the null hypothesis at 5% significance level. The critical value for $F_{2,\infty}$ at 5% significance level is 3.
- Ex. If H_0: $\beta_1 = \beta_2 = 0$ and F-stat >3, we reject the null and conclude that *at least* one of the coefficients is not zero.

Nonlinear Regression (1)

- Polynomials in X

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

- The coefficients do not have a simple interpretation because it is impossible to change X holding X^2 & X^3 constant.

Q. Which polynomial model to choose?

1. Linear vs. Quadratic? T-test on H_0: $\beta_2 = 0$
2. Quadratic vs. Cubic? T-test on H_0: $\beta_3 = 0$
3. Linear vs. non-linear? F-test on H_0: $\beta_2 = \beta_3 = 0$

If fail to reject H_0: $\beta_2 = \beta_3 = 0$, linear model is preferred. If reject, at least one of the coefficients is not zero.

Nonlinear Regression (2)

- Log transformation
- If $\frac{\Delta X}{X}$ is small, $ln(X + \Delta X) - lnX = ln\frac{X+\Delta X}{X} = \ln(1 + \frac{\Delta X}{X}) \cong \frac{\Delta X}{X}$

Case	Regression Specification	Interpretation of β_1
Linear-Log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$	1% change in X➔ 0.01 β_1 change in Y
Log-Linear	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$	1 unit change in X➔100 β_1 % change in Y
Log-Log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$	1% change in X➔ β_1 % change in Y

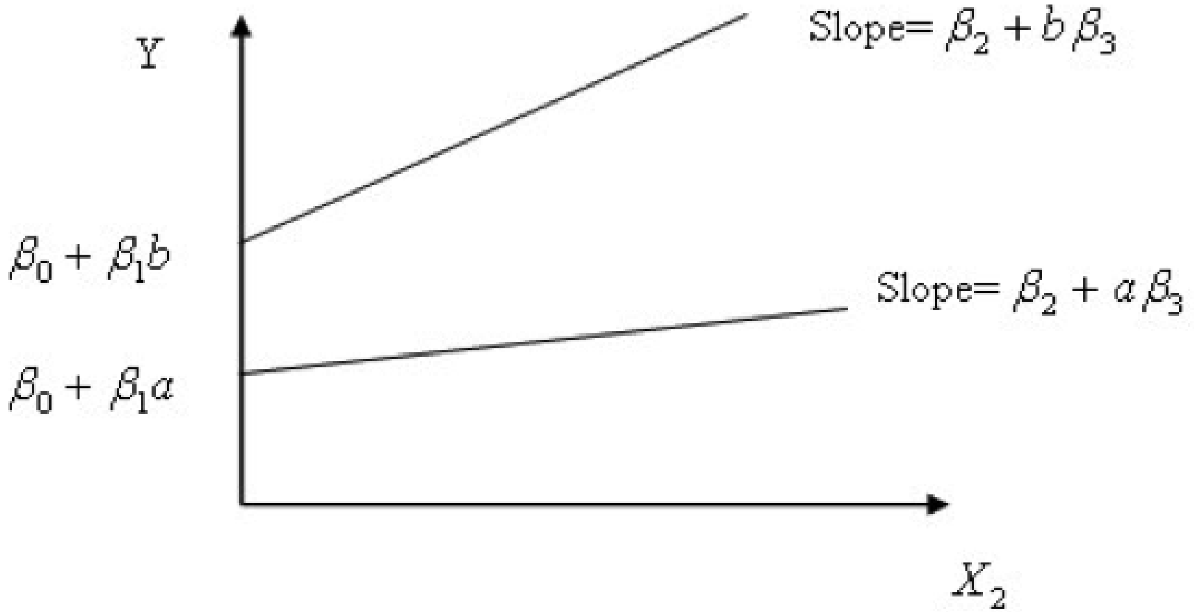
- In log-log specification, β_1 has elasticity implication.

Nonlinear Regression (3)

- The interactions
- Consider regression $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$
- If we believe that the effect of X_{2i} on Y depends on X_{1i}, we can include the "interaction term" $X_{1i}*X_{2i}$ as a regressor;

  $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} * X_{2i}) + u_i$
- Suppose we want to plot the regression line of Y on X_{2i} when X_{1i}=a and X_{1i}=b. Assume that all coefficients are positive and a and b are also positive with a<b.
- When X_1 = a, $Y_i = \beta_0 + \beta_1 a + (\beta_2 + \beta_3 a) X_{2i}$
- When X_1 = b, $Y_i = \beta_0 + \beta_1 b + (\beta_2 + \beta_3 b) X_{2i}$

Interaction terms



- The intercept of the bottom regression line represents the predicted value of Y when X_2= 0 when X_1=a .
- The slope of the bottom regression line represents the effect on Y of a unit change in X_2 when X_1=a .

Example of the Interaction Terms

- Let Black=1 if observation is black and 0 otherwise. Exp= years of experience. If we believe that the effect of experience on wage depends on individual's race, we can add the interaction term, Black *Exp, to the regression.

$$lwage_i = \beta_0 + \beta_{black} Black + \beta_{\exp} Exp + \beta_{black\_\exp} (Black * Exp) + u_i$$

- If Black=1, $Y_i = \beta_0 + \beta_{black} + (\beta_{\exp} + \beta_{black\_\exp}) Exp + u_i$
- If Black=0, $Y_i = \beta_0 + \beta_{\exp} Exp + u_i$
- To see whether the average wage for no experience (intercept) is the same for two groups, test $H_o : \beta_{black} = 0$
- To see whether effect of experience on wage (slope) is the same for two groups, test $H_0 : \beta_{black\_\exp} = 0$

Models for Binary Dependent Variable

- Linear Regression Model
  - Linear Probability Model
- Non-Linear Regression Models
  - Probit Model
  - Logit Model
  - Ordered Probit Model
- The main idea of the model with a binary dependent variable is to interpret the population regression as the probability of success given X; $Pr(Y = 1 | X)$.
- If Y is binary,
  $E(Y|X)=1*Pr(Y=1|X)+0*Pr(Y=0|X)=Pr(Y=1|X)$

Linear Probability Model (LPM)

- $Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$
  $E(Y_i \mid X_i) = \Pr(Y_i = 1 \mid X_{1i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}$
- This is a probability model. β_1 tells you the change in probability that Y=1 for a unit change in X_1, holding other regressors constant.
- The predicted value from the regression is the predicted probability that Y=1 for the given values of the regressors.
- Ex. For regression model $Y_i = \beta_o + \beta_1 X_i + u_i$
  $E(Y_i \mid X = 5) = \Pr(Y_i = 1 \mid X = 5) = \hat{\beta}_o + \hat{\beta}_1 * 5$

Advantage and Disadvantages of LPM

- Advantage:
  - Simple to estimate and to interpret. Statistical inference is the same as for multivariate regression model.
  - One unit increase in X increases Pr(Y=1) by 100*β_1 *percentage point*.
- Disadvantages:
  - The predicted probabilities can be < 0 or >1, which does not make sense in probability.
  - This can be solved by introducing nonlinear probability model: probit and logit regression

Probit Regression Model

- Probit regression models the probability that Y=1 using cumulative standard normal distribution, $\Phi(z) = P(Z \le z)$ where $Z \sim N(0,1)$.
- $E(Y_i \mid X_1, X_2) = \Pr(Y_i = 1 \mid X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \Phi(z)$
- Predicted probability of Y=1 given X is calculated by computing the "z-score", $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, and looking up this z-value in the standard normal distribution table.
- β_1 measures the change in the z-score for a unit change in X_1 holding X_2 constant, not the change in the predicted probability of Y=1 given X, $\Pr(Y = 1 \mid X_1, X_2)$.

Logit Regression Model

- Logit regression models the probability that Y=1 using the logistic function evaluated at $z = \beta_0 + \beta_1 X_1$

$$E(Y_i \mid X_i) = \Pr(Y_i = 1 \mid X_i) = F(\beta_0 + \beta_1 X_1) = \frac{1}{1 + \exp - (\beta_0 + \beta_1 X_1)}$$

- Predicted probability of Y=1 given X is calculated by computing z-value, $z = \beta_0 + \beta_1 X_1$, and looking up this z-value in the standard logistic distribution table
- β_1 measures the <u>change in the z-score</u> for a unit change in X_1, not the change in $\Pr(Y = 1 \mid X_1)$.

Issues in Probit/Logit Model

- The "S" shape of the curves of probit and logit models ensures:
  1. $0 \leq \Pr(Y = 1 | X) \leq 1$, for all X
  2. $\Pr(Y = 1 | X)$ to be increasing in X for positive coefficients.
- We can interpret the *sign* of the coefficients in the usual way, but we cannot interpret the *magnitude* of the coefficients in the usual way. The coefficient itself is meaningless.
- To compute the effect on the probability of Y=1 of change from X_1=a to X_1=b, holding other regressors constant (say X_2=c), calculate "before-and-after" change;

$$\Pr(Y_i = 1 | X_1 = b, X_2 = c) - \Pr(Y_i = 1 | X_1 = a, X_2 = c)$$
$$= \Phi(\beta_0 + \beta_1 * b + \beta_2 * c) - \Phi(\beta_0 + \beta_1 * a + \beta_2 * c)$$

Ordered Probit Model

- Ordered probit model is used when the dependent variable consists of ordered categories. For example,
  - Y_i=1 if $Y_i^* \le$ cutoff1
  - Y_i=2 if cutoff1 < $Y_i^* \le$ cutoff2
  - Y_i=3 if cutoff2 < Y_i^*

  where Y_i^* is the latent variable (usually true values).
- Suppose $Y_i^* = \beta_0 + \beta_1 X_i + u_i$ where $u_i \sim N(0,1)$.

  $\Pr(Y_i = 1 \mid X) = \Phi[cutoff\,1 - (\beta_0 + \beta_1 X_i)]$

  $\Pr(Y_i = 2 \mid X) = \Phi[cutoff\,2 - (\beta_0 + \beta_1 X_i)] - \Phi[cutoff\,1 - (\beta_0 + \beta_1 X_i)]$

  $\Pr(Y_i = 3 \mid X) = 1 - \Phi[cutoff\,2 - (\beta_0 + \beta_1 X_i)]$
- To find the effect of changing X from a to b on probability of being in category y, compute the "before-and-after" change: $\Pr(Y_i = y \mid X = b) - \Pr(Y_i = y \mid X = a)$

Example of Probit Model

- Suppose we regressed mortgage denial rate on race and the ratio of debt payments to income (P/I), and obtained the following result.
- Deny= -2.26 + 2.74*(P/I) + 0.71*black

  (.16) (.44) (.08)
- $\Pr(deny=1 \mid P/I, black) = \Phi(-2.26 + 2.74 \times P/I + 0.71 \times black)$

Q. Calculate the estimated effect of being black when P/I = 0.3.

  - $\Pr(deny=1 \mid 0.3, black=1) = \Phi(-2.26+2.74 \times .3+.71 \times 1) = .233$
  - $\Pr(deny=1 \mid 0.3, black=0) = \Phi(-2.26+2.74 \times .3+.71 \times 0) = .075$
  - $\Pr(deny=1 \mid 0.3, black=1) - \Pr(deny=1 \mid 0.3, black=0) = .158$

- The rejection probability increases by .158. Thus, the chance of being denied on mortgage is 15.8 *percentage point* higher for the black applicants compared to the non-black applicants.

Types of Data

- Panel Data: N different entities are observed at T different time periods. If the N entities are the same for each period, it is called *longitudinal data*.
  1. Balanced Panel: All variables are observed for each entity and each time period.
  2. Unbalanced Panel: There are missing data for at least one time period for at least one entity.
- Cross-Sectional Data: N different entities are observed at the same point in time
- Time Series Data: 1 entity is observed at T different time periods

Panel Data

- Since each entity is observed multiple times, we can use fixed effects to get rid of the OVB, which results from the omitted variables that are invariant within an entity or within a period.
- Entity Fixed Effects control for omitted variables that are constant within the entity and do not vary over time
  - ex. gender, race, or cultural and religious characteristics of each state
- Time Fixed Effects control for omitted variables that are constant across entities but vary over time
  - ex. national level anti-crime policy or national safety standard each year

Fixed Effects Regression Model

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \ldots + \beta_k X_{kit} + (\gamma_1 D_1 + \ldots + \gamma_{N-1} D_{N-1}) + (\delta_1 P_1 + \ldots + \delta_{T-1} P_{T-1}) + u_{it}$$

$$Y_{it} = \beta_1 X_{1it} + \ldots + \beta_k X_{kit} + \gamma_i + \delta_t + u_{it}$$

where i=1, 2, …,N and t=1, 2, ...,T.

- # of observation = i * t
- X_{kit} denotes the kth variable for individual i at time t
- $D_1, D_2, \ldots, D_{N-1}$ are entity dummy variables
- $P_1, P_2, \ldots, P_{T-1}$ and time dummy variables
- γ_i are entity fixed effect, and δ_t are time fixed effect.
- γ_i and δ_t is unknown intercept to be estimated for each entity and time period.

Special case: FE Regression and Difference Regression for T=2

- Regression with entity FE:

(1) $Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$ for t=1,2

- Change regression for entity FE regression:

(2) $Y_{i2} - Y_{i1} = \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$

  - β_1 in (1) = β_1 in (2) without intercept.

- Regression with entity FE and time FE:

(3) $Y_{it} = \alpha_i + \beta_1 X_{it} + \delta P_{2i} + u_{it}$ for t=1,2

- Change regression for entity FE and time FE regression.

(4) $Y_{i2} - Y_{i1} = \delta + \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$

  - β_1 in (3) = β_1 in (4) with intercept.

Differences-in-Differences Estimator

- $Y_{it} = a + \alpha Treatment_{it} + \beta After_{it} + \gamma (Treatment_{it} * After_{it}) + u_{it}$
  - Treatment= 1 for treatment group, 0 for control group
  - After= 1 after the program, 0 before the program
- For treatment group,
  - After the program: $Y = a + \alpha + \beta + \gamma$
  - Before the program: $Y = a + \alpha$
  - The effect of the program: After-before: $\beta + \gamma$
- For control group,
  - After the program: $Y = a + \beta$
  - Before the program: $Y = a$
  - The effect of the program: After-before: β
- Differences-in-Differences estimator: $(\beta + \gamma) - \beta = \gamma$

Serial Correlation in Panel Data

- If there is a correlation of the error terms from year to year for a given entity, called serial correlation, conditional on the regressors and entity FE, the OLS estimate of β in panel data is still *unbiased*.
- However, the estimated standard errors will be wrong, which implies that we can't carry out the usual hypothesis tests. In this case, we have to use Heteroskedasticity-Autocorrelation-Consistent (HAC) standard errors.
- Clustered HAC SE is robust to both heteroskedasticity and serial correlation of the error terms in panel data. It allows the errors to be correlated within a cluster, but not across the clusters.

Internal and External Validity

- Threats to internal validity
  1. OVB
  2. Misspecification of the Functional Form
  3. Measurement Error
  4. Sample Selection Bias
  5. Simultaneous Causality Bias
     - $E(u|X)\neq 0$, so OLS estimates are biased.
- Threats to external validity
  1. Differences in population studied and population of interest
  2. Difference in setting studied and setting of interest

Endogenous independent variable

- The OLS assumption of exogeneity, $E(u|X) = 0$, is violated if there is correlation between X and u. Then, X is said to be endogenous variable.
- If the exogeneity assumption is violated, the coefficient of X in OLS regression is biased and inconsistent.
- IV regression helps us get an unbiased and consistent estimator of the causal effect of changes in X on Y.

Q. When should you use IV regression?

1. When you believe that X is correlated with error term u (X is endogenous) AND
2. You have valid instruments.

Instrumental Variable (IV) Regression

- Consider regression $Y_i = \beta_0 + \beta_1 X_i + u_i$
- Suppose E(u|X) ≠ 0; X is endogenous.
- IV regression breaks X into two parts by using the instrumental variable Z
  1. a part that is correlated with u
  2. a part that is not correlated with u (this part is correlated with Z).
- By focusing on the variation in X that is not correlated with u using the help from the instrument Z, it is possible to obtain an unbiased and consistent estimate of β_1.

Conditions for a Valid IV

- For Z to be a valid instrument, it must satisfy two conditions:

1. Instrument relevance: $Corr(Z, X) \neq 0$
   - The instrument Z should be correlated with X since we will use this correlation to take out the part of X, which is uncorrelated with the regression error term.
2. Instrument exogeneity: $Corr(Z, u) = 0$
   - Z must be exogenous. Z should affect Y only through X, not through u_i. In other words, Z has an indirect effect on Y through X, but has no direct effect on Y. Z is correlated with X but uncorrelated with any other variables that can affect Y.

Two Stage Least Squares (TSLS)

- Suppose we have a model: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$ where X is endogenous variable and W is included exogenous variable.
- First stage regression : $X_i = \pi_0 + \pi_1 Z_i + \pi_2 W_i + v_i$
  - Regress X_i on Z_i and W_i to isolate the part of X_i that is uncorrelated with u_i.
  - Compute the predicted values of X_i: $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i + \hat{\pi}_2 W_i$
  - Now, $\hat{X}_i$ is uncorrelated with u_i by construction.
- A regression is called *reduced form* if it regresses an endogenous variable on all the available exogenous variables (both Zs & Ws). The first stage regression and the regression of Y on Zs & Ws are in reduced form.

TSLS (cont.)

- Second stage regression : $Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 W_i + u_i$
  - Regress Y_i on $\hat{X}_i$ and W, using OLS.
  - Because $\hat{X}_i$ is uncorrelated with u_i, all the regressors are exogenous in the second stage regression.
  - The resulting estimator of β_1 is called TSLS estimator.
  - With some algebra, we can show that;

$$\hat{\beta}_1^{2SLS} = \frac{Cov(\hat{X}_i, Y_i)}{Var(\hat{X}_i)} = \frac{\sigma(Z,Y)}{\sigma(Z,X)} = \frac{\alpha}{\delta}$$

where σ is the sample covariance, α is the coefficient of Z in the regression of Y on Z, and δ is the coefficient of Z in the regression of X on Z.

General IV Regression Model

$$Y = \beta_0 + \beta_1 X + \ldots + \beta_k X_k + \beta_{k+1} W_1 + \ldots + \beta_{k+r} W_r + u$$

- $X_1, \ldots, X_k$:k endogenous regressors (correlated with u)
- $W_1, \ldots, W_r$:r included exogenous regressors (uncorrelated with u)
- $Z_1, \ldots, Z_m$: m instrumental variables
- The coefficients $\beta_1, \ldots, \beta_k$ are said to be:
  1. Exactly identified if m = k
     - The number of instruments = the number of endogenous regressors.
  2. Over-identified if m > k
     - The number of instruments > the number of endogenous regressors.
  3. Under-identified if m < k
     - The number of instruments < the number of endogenous regressors. In this case, you need to find more instruments.

Testing the Relevance of Instruments

Step 1: Run the <u>first stage</u> regression with all Zs and Ws;

$$X_i = \pi_0 + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \ldots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \ldots + \pi_{m+r} W_{ri} + v_i$$

Step 2: Run the joint hypothesis to test H_0: $\pi_1 = \pi_2 = \ldots = \pi_m = 0$

Step 3: The instruments are relevant if at least one of the m coefficients is nonzero. <u>If the F-statistic >**10**, reject the null</u> and conclude that our instrument(s) are highly <u>relevant</u>. If the F-statistic < **10**, our instruments are weak.

- Weak instruments explain very little of the variation in *X*. If instruments are weak, the sampling distribution of TSLS estimator and its *t*-statistic are not normally distributed even with large sample, so $\hat{\beta}_1^{2SLS}$ is inconsistent.

Testing the Exogeneity of Instruments

- We use the **J-test** if the model is over-identified (m>k);
  If instruments are exogenous, they are uncorrelated with u_i, and also (approximately) uncorrelated with the residual from TSLS regression.

Step 1: Run the TSLS regression to find the residuals $\hat{u}_i^{TSLS}$.

Step 2: Regress the residuals on the Zs and Ws;

$$\hat{u}_i^{TSLS} = \delta_0 + \delta_1 Z_{1i} + \ldots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \ldots + \delta_{m+r} W_{ri} + \varepsilon_i$$

Step 3: Compute the F-statistic testing that all instruments are exogenous. H_0: $\delta_{1i} = \delta_{2i} = \ldots = \delta_{mi} = 0$

Step 4: Compute the *J*-statistic = m*F. J-statistic has χ^2 distribution with m (number of instruments) – k (number of endogenous variables) degrees of freedom; $J \sim \chi^2_{m-k}$

Step 5: If the J-statistic < the critical value, then we fail to reject the null and conclude that the instruments are exogenous.

Time Series Data

- Data for the same entity observed over time. There is one observation at one point in time.
- When we analyze time series data, we are mainly interested in "impulse(shock)-response" relationship. For example, we can examine the impact of rise in oil price on stock market index or the impact of monetary policy on aggregate demand.
- The purpose in time series study is often to make a good forecast.
- The first lag of Y_t is Y_{t-1}, the second lag is Y_{t-2}, etc.
- The correlation of a series with its own lagged values is called <u>autocorrelation</u> (or serial correlation).

Stationarity

- Time series Y_t is <u>stationary</u> if its distribution does not change over time; $E(Y_t)$ and $Var(Y_t)$ is constant over time, and the $Cov(Y_t, Y_{t-j})$ does not depend on t.
- Stationarity implies that Y_t has some trend, so history is relevant for forecasting the future.
- If Y_t is not stationary, our normal regression theory breaks down: The *t*-statistics don't have standard normal distributions even in a large sample, so we can't check significance of the coefficients in the usual way.
- If data is not stationary, we can transform it usually by taking the <u>first differences</u> to make it stationary before estimation.

Forecasting Model (1)

- **Autoregressive Model (AR):**
  The p^{th}-order autoregression, AR(p), is a linear regression in which Y_t is regressed on its first p lagged values:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t$$

  - Assumption 1: $E(u_t | Y_{t-1}, Y_{t-2}, \ldots) = 0$: The error terms are serially uncorrelated.
  - Assumption 2: Y_t needs to be stationary.
  - The best forecast of Y_{T+1} based on its entire history depends on only the most recent p past values:

$$Y_{T+1|T} = \beta_0 + \beta_1 Y_T + \beta_2 Y_{T-1} + \ldots + \beta_p Y_{T-p+1}$$

  - The coefficients do <u>not</u> have a causal interpretation.

Forecasting Model (2)

- **Autoregressive Distributed Lag (ADL) Model:**

  The autoregressive distributed lag model, ADL(p,r), is a linear regression in which Y_t is regressed on p lags of Y and r lags of X:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \ldots + \delta_r X_{t-r} + u_t$$

  - Assumption 1: $E(u_t \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}, X_{t-1}, X_{t-2}, \ldots, X_{t-r}) = 0$.

  No serial correlation and no additional lags beyond p and r belong to ADL(p,r) model

  - Assumption 2: The random variable (Y_t, X_t) are stationary.
  - The coefficients do <u>not</u> have a causal interpretation.

Model Selection

- Tradeoff: If including too few lags, the coefficients could be biased. If adding too many lags, forecast errors could be too large.
- When we decide the optimal length of lag (p), use Information Criterion: BIC or AIC
- Bayesian Information Criterion: $BIC = \ln\left[\frac{RSS(p)}{T}\right] + (p+1)\frac{\ln T}{T}$
  - First term: Always decreasing in p (larger p, better fit)
  - Second term: Always increasing in p. (Penalty term for increased forecast error due to increased p)
  - ➢ Choose p that minimizes BIC.

Dynamic Causal Effects

- <u>Dynamic causal effect:</u> effect of a change in X on Y over time. A shock to X_{t-2} affects Y_t directly and indirectly through X_{t-1}.
- **The Distributed Lag model (DL):**

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \ldots + \beta_{r+1} X_{t-r} + u_t$$

  - β_1 = <u>impact effect of a change in X</u> = effect of a change in X_t on Y_t, holding past X_ts constant
  - β_2 = <u>1-period dynamic multiplier</u> = effect of a change in X_{t-1} on Y_t, holding constant X_t, X_{t-2}, X_{t-3},…
  - β_3 = <u>2-period dynamic multiplier</u> = effect of a change in X_{t-2} on Y_t, holding constant X_t, X_{t-1}, X_{t-3},…
  - <u>Cumulative dynamic multipliers</u>= the 2-period cumulative dynamic multiplier = $\beta_1 + \beta_2 + \beta_3$

Serial Correlation and HAC SE

- When u_t is serially correlated, OLS coefficients are consistent, but the usual SE without correcting for serial correlation is wrong. So t-statistic based on this SE is also wrong.
- In panel data, we use clustered HAC SE to correct for serial correlation. It allows the errors to be correlated within a cluster but not across the clusters. This approach requires the number of entities to be > 1.
- In time series data, the number of entities $= 1$ for each period, so clustering is not available. Instead, we use <u>Newey-West HAC SE</u>, which estimates correlations between lagged values in time series.

Newey-West HAC SE

- Too little lags: some of the autocorrelation that might be important for estimation could be missing from the regression, so the estimator could be biased.
- Too many lags: there could be a large estimation error
- The rule of thumb to decide how many lags we will have to estimate is to use the truncation parameter m (number of lags) to employ Newey-West HAC SE: $\underline{m=0.75*T^{1/3}}$, where T is the number of observation (or the number of periods)
- In AR(p) model or ADL(p,q) model, the error terms are serially *uncorrelated* if we include enough lags. Thus, *no need* to use Newey-West HAC SE if the model includes the optimal number of lags. In DL model, however, serial correlation is unavoidable and Newey-West HAC SE is preferred.

Good Luck on Your Exam!!!