ECON 1123 Final Exam Cheatsheet

April 29, 2023

1 Statistics

- Variance of a sum:
$$Var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab \cdot cov(X, Y)$$
- Significance: A measure is significant at the 5% level if:
  - The test statistic is larger than the 5% critical value for the test. OR
  - The p-value is less than .05 OR
  - The 95% confidence interval does not include 0.
- 95 % Confidence Interval:
  - An interval that will contain the true value of a point estimate in 95% of repeated samples.
  - A set of values that cannot be rejected at a 5% level with the data we have.

2 Interpretations

Linear-Linear Regression	A 1 unit change in x is associated with an alpha_1 unit change in y
Log-Linear Regression	A 1 unit change in x is associated with a 100alpha_1% change in y
Linear-Log Regression	A 1% change in x is associated with a 0.01alpha_1 unit change in y
Log-Log Regression	A 1% change in x is associated with an alpha_1% change in y

3 Omitted Variable Bias

$$Y = \alpha_0 + \alpha_1 X \text{ (Short Form)}$$
$$Y = \beta_0 + \beta_1 X + \beta_2 W \text{ (Long Form)}$$
$$W = \gamma_0 + \gamma_1 X$$

- Omitted variable must be something correlated with X and Y, but should not be part of a causal channel.
- Steps:
  1. Choose a valid omitted variable
  2. Sign β_2 (correlation between omitted variable and Y)
  3. Sign γ_1 (correlation between omitted variable and X)
  4. Use the above two to sign the bias ($bias = \beta_2 \times \gamma_1$)
  5. Use the above to decide if we have overstated or understated the causal effect.
- HINT: if α_1 and $bias$ have the same signs, you have overstaded the causal effect. If they have different signs, you have understaded causal effect.

4 Threats to Validity

- Internal: Have we correctly estimated coefficients and standard errors within our sample?
  - Omitted Variable Bias
  - Functional Form Bias
  - Measurement Error (in X) (Attenuation Bias)
  - Bad Controls
  - Sample Selection Bias
  - Simultaneous Causality Bias
  - Wrong Standard Errors
    All of the above except the last one are a result of conditional mean independence being violated, meaning after controls, the variable of interest is as good as randomly assigned.
- External: Can our results be extended to other populations?
  - Different place, time, group of people, etc.

5 Standard Errors

- Heteroskedasticity-Robust vs. Pooled: Pooled standard errors only work when variance is constant across our dataset. HR standard errors work either way.
- Clustered SEs: We use clustered standard errors to account for serial correlation in the data. This is necessary when data points are not independent of each other, as in panel data. We always cluster at the level of policy change (change in treatment).
- Newey-West SEs: Newey-West standard errors account for autocorrelation in the error terms. You don't need to use these if you're already accounting for autocorrelation in your regression!

6 Instrumental Variables

- Instrument Z, Treatment X, outcome Y
- First-stage: $X = \pi + \pi_1 Z$
- Reduced-Form: $Y = \alpha_0 + \alpha_1 Z$
- $\beta_1 = \frac{\alpha_1}{\pi_1}$
- Requirements for Valid Instrument:
  - Relevance: Z plausibly has impact on X. We measure this with first-stage f-statistic. If there is just one instrument, we can calculate this using the first-stage regression:

    $$F = \left(\frac{\pi_1}{SE(\pi_1)} \right)^2$$

    and the critical value is 23.1
  - Exogeneity:
    - As good as randomly assigned: With any necessary controls, the instrument is as good as randomly assigned.
    - Instrument only effects outcome through treatment. Instrument cannot affect outcome directly and cannot affect it through another variable.
- LATE vs. ATE:
  - Local average treatment effect is measured by 2SLS and is a weighted average where compliers have greater weights than non-compliers.
  - Compliers (grasshoppers) are those for whom the instrument affects the treatment. Non-compliers (ants) are those for whome the instrument does not affect the treatment.

– Steps to LATE vs. ATE Problem:

1. Eliminate 3 cases where LATE = ATE.
   (a) π_i is constant, meaning compliance levels are the same for everyone.
   (b) β_1 is constant, meaning the treatment effect is the same for everyone.
   (c) $cov(\pi_1, \beta_1) = 0$, meaning being a complier or non-complier is not related to the treatment effect in any way.
2. Identify who the compliers and non-compliers are. (These can be a smaller group within compliers or non compliers, eg. you could say rich people are more or less likely to be compliers)
3. State whether the treatment effect (effect of X on Y) is larger (in absolute value) for compliers or non-compliers. Make sure you are ignoring the instrument here, and only looking at the effect of X on Y.
4. Finally, use the above to state whether LATE is bigger than or smaller than ATE.

7 Cumulative Impluse Response Functions

- For a regression of the form:
$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{it-1} + \beta_3 X_{it-2}$$
- Where $\Delta X_{it} = X_{it} - X_{it-1}$
- Then for the additional regression:
$$Y_{it} = \delta_0 + \delta_1 \Delta X_{it} + \delta_2 \Delta X_{it-1} + \delta_3 X_{it-2}$$
- $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, and $\delta_3 = \beta_1 + \beta_2 + \beta_3$
- Long-Run Cumulative Effect: $\beta_1 + \beta_2 + ...$, or δ_n.
- Cumulative Impulse Response Function. Graphing number of lags on the x axis, and total cumulative effect to that point on the y.