

EC420 Final Review

Kirkpatrick

4/20/2021

Final Review

I am giving you an outline of what we will talk through during our final class which will serve as a review session. Our final exam is scheduled for **Thursday, April 29 5:45pm - 7:45pm** on D2L. It will be much like the Midterm, but with a longer time period – 120 minutes vs. 90 for the midterm. I will likely *not* scale up the number of questions by $120/90 = 1\frac{1}{3}$. I will have usual office hours the Tuesday of finals week, and am available for office hours outside of those hours for those of you who have conflicts. Please use this resource.

Part 1 (before the midterm)

The emphasis on the material from before the midterm will be on topics and concepts that are used in the post-midterm half of the class.

Stats Review

- Population vs. Sample Statistics
 - Greek letters like μ for population
 - Non-greek letters like \bar{x} for sample
- Expectation (mean) of a Random Variable
 - Operations with the Expectation
 - * $E[aX + Y] = aE[X] + E[Y]$
- Variance of a Random Variable
 - Formulat for Variance
 - * $Var(X) = \sum_{n=1}^N (x_n - \mu_X)^2$
 - Operations with Variance
 - * $Var(aX + Y) = a^2Var(X) + Var(Y) + 2Cov(X, Y)$
 - Variance is a measure of *dispersion*
- PDFs and CDFs
- Normal distribution
- Sum of normals is normals
- Φ (std. normal CDF) and ϕ (std. normal pdf)

- Binary, Count, and Continuous distributions
- Correlation and Covariance
 - Formula for covariance
 - * $Cov(X, Y) = \sum_{n=1}^N (x_n - \mu_X)(y_n - \mu_Y)$
- Independence and Mean Independence
- Estimator
 - The plan for generating an estimate
- Estimate
 - The result of an estimator once it is taken to the data
- Unbiased
 - $E[\hat{\theta}] = \theta$ if $\hat{\theta}$ is an unbiased estimator
- Consistent
 - As $n \rightarrow \infty$, a consistent estimator is one that gets closer and closer to the true estimate
 - $\frac{1}{N} \sum x_i$ is a consistent estimator for μ_X .
- Sample mean, variance
 - Sample variance: $\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2$
 - Degrees of freedom
- Sample distribution of a statistic

Linear Regression (single, multiple)

- Population Regression Function (PRF)
- Sample Regression Function (SRF)
- Calculating $\hat{\beta}$ in a single variable
 - $\frac{Cov(X, Y)}{Var(X, Y)}$
 - Rescaling X by a constant c
 - Rescaling Y by a constant c
- Calculating \hat{y} using $\hat{\beta}$
 - And $u = y - \hat{y}$
- Getting a sample estimate of $\sigma_u^2, \hat{\sigma}_u^2$
 - Degrees of freedom: $N - K - 1$
- And finally, calculating $\hat{se}(\hat{\beta})$ using $\hat{\sigma}_u^2$ and SST_x
- Interpreting coefficients
 - Slope of the line of best fit

- $\frac{dY}{dX}$
- Hypothesis testing coefficients
 - * $H_0 : \hat{\beta} = 0$
 - * $H_A : \hat{\beta} \neq 0$
 - * Constructing a t-statistic
 - P-value
 - 95% Confidence Interval
- R^2
 - Interpreting
 - Calculating
 - * $SST = SSR + SSE$
- Multiple regression
 - Partialling out x_2 from x_1
 - Calculating $\hat{se}(\hat{\beta}_j)$, the std. error of the estimate in multiple regression
 - Interpretation of β_j “Change in $E[Y]$ per one-unit increase in X , ceteris paribus”
- Issues with multiple regression
 - Perfect multicollinearity (fatal to regression)
 - Multicollinearity (potential problem)
- F-test
 - Tests whether or not the full model is any better than a model with no X ’s (just a β_0)
 - Joint tests:
 - * Are a group of coefficients *jointly* equal to zero?
- Functional form
 - Log transformation interprets as percent change
 - Polynomials for flexible functional form
- Binary X variables
 - Indicator function
 - Interpretation
 - * Intercept shift for group
- Fixed Effects
 - Account for what is common across a group or time
 - Constructed as set of dummies
 - * One “base level” value
 - *Interpret* fixed effect
- Interactions
 - Binary or Dummy interacted with continuous
 - * Slope shift instead of intercept shift
 - Continuous x continuous interaction
- MLR1-MLR6
 - Relaxing MLR6 to use asymptotic properties
 - Relaxing MLR5 for heteroskedasticity

Counterfactuals, Selection Bias, Experiments and Randomization

- Counterfactual
 - Treatment
 - Potential outcomes
 - Only observe one of the potential outcomes
 - Counterfactual is the non-observed potential outcome -Average Treatment Effect (ATE)
 - $E[Y_1 - Y_0]$
 - Rubin Potential Outcomes framework
 - * Everyone has two possible outcomes
 - * (Y_{i1}, Y_{i0})
- Selection Bias
 - If treatment, D is not independent of (Y_{i1}, Y_{i0}) , then there is selection bias.
 - This also means $E[U|X] \neq 0$
 - Showed naive comparison of means gets $SATE$ + selection bias
 - * Selection Bias: $E[Y_0|D = 1] - E[Y_0|D = 0]$
- Treatment Effects
 - A variable of interest, D , potentially with covariates X
 - Is D randomly assigned?
 - Randomization removes selection bias
 - * But not always possible
 - How do we get an unbiased estimated of β_D when D is endogenous
 - * $E[U|D] \neq 0$

Instrumental Variables

- When D is endogenous
 - Determined within the system
- One solution is an instrument for D
 - Z is an instrument if:
 - * Relevant First Stage $Z \rightarrow D$
 - * Exogenously (or as good as randomly) assigned
 - * Z affects Y *only* through it's affect on D (the *exclusion restriction*)
- Use 2SLS to estimate:
 - $D = \delta_0 + \delta_1 Z_1 + \delta_2 x_1$
 - \hat{D}
 - $Y = \beta_0 + \beta_1 \hat{D} + \beta_2 x_1$
 - * Where x is exogenous, z is instrument, D is endogenous, and Y is outcome.
- Testing for relevant first stage
- Overidentification
- Local Average Treatment Effect
 - $LATE$
 - Compliers, defiers, always-takers, never-takers and the assumed-to-not-exist *defiers*
- Simultaneous equations
 - Two equations where the variable of interest affects and is affected by the outcome
 - Structural when the equations reflect an economic relationship like “supply” and “demand”
 - * Supply and demand of labor means the wage is the equilibrium of supply intersecting demand
 - Need at least one instrument per endogenous variable in each equation
 - Know what can and cannot be estimated in a simultaneous model

Diff in Diff

- Treated group vs. control group
- Panel data
- Endogeneity may be because the treatment is correlated with something unobserved about the treated units
- Diff in Diff as a counterfactual
 - Identifying assumption: **Parallel Trends Assumption**
 - If we can assume that the treatment would have followed the same *trend* as the control group, then we have a counterfactual
- Control for treatment/control group with fixed effects
- Control for trends with time fixed effects
- Implementation
 - basic: $Y = \beta_0 + \beta_1 * 1(Treatment) + \beta_2 * 1(Post) + \beta_3 1(Treatment) * 1(Post)$
 - More flex: $Y = \beta_0 + \beta_1 1(Treatment) * 1(Post) + \Gamma_i + \theta_t$
- Interpretation
- Hypothesis testing
 - Same tools as multiple regression
- DiD estimator *is* just a coefficient in regression if assumptions hold.
 - And is *ATE* under those assumptions.

Regression Discontinuity

- D might be endogenous (SAT example)
- DID not useful, no instrument
- Ingredients:
 - Running variable
 - Threshold
 - Window
- Identifying Assumption: within a small enough window around the threshold, treatment is as good as randomly assigned
- Implementing in a regression
 - Smooth control for running variable
 - Dummy for passing threshold
- Interpretation
- Testing assumptions
 - Do the other variables have an effect as they cross the threshold?
 - * Hopefully no.
- Fuzzy RD
 - Not always perfect compliance at threshold
 - Instrument treatment using threshold

Time Series (As promised, will not be on Final)

- *Concept* of observing only a chain of events in a sample
 - Having 100 observations is meaningless if they are strongly correlated
- Strict Exogeneity necessary for standard MLR assumptions
 - Why is this hard to assume?
 - $E[u|\mathbf{X}] = 0$
- Stationarity
 - Joint distribution is unchanged with t
- Weak dependence
 - Correlation dies out eventually over time
- TSR1' to TSR5'
 - Need weak dependence, contemporaneous exogeneity $E[u_t|\mathbf{x}_t] = 0$
 - And need “no serial correlation”
 - Then get consistent estimator for β !
- FDL, MA(1), MA(q), AR(1), AR(q)

```
knitr::knit_exit()
```