

Week 16: Final Exam Review and Reading/Evaluating Design-based Papers in the Future

JJ Chen

May 1, 2015

Today's Plan

- ▶ Practice final exam questions
- ▶ Reading/Evaluating design-based papers in the future

Preparation

- ▶ Most of the time:
 - ▶ Some kinds of regressions or conditional mean comparisons + ID assumptions \rightsquigarrow causal parameter of interest
- ▶ Be familiar with common ID strategies
 - ▶ *Describe ID Asm in words for: thought experiment, OLS, Matching, IV, RD, DID, and all their friends
 - ▶ **Think about implications/tests of ID Asm: if the ID Asm is correct, then we should be able to observe XYZ
 - ▶ **Write down the exact math/stat definitions for ID Asm
- ▶ Know how to write down a regression
 - ▶ *Try to be familiar with all the regressions in MHE; **OVb and Reg Anatomy**
 - ▶ ***Read abstracts of assigned papers, once you understand the source of ID, try to write down a regression and compare

Practice Final Exam Questions I

T or F: Including FE in a regression always reduces an OVB because FE control for unobserved heterogeneity.

- ▶ Consider

$$Y_{it} = \alpha + \rho D_{it} + \beta_i OV1_i + \gamma_t OV2_t + \delta_{it} OV3_{it} + \eta_{it}$$

- ▶ How does individual FE controls for OV1? What about OV2?
 - ▶ OV1 doesn't change over time, so any change in Y over time cannot be caused by OV1
- ▶ When we have a panel or similar data structure, the source of variations to identify ρ can be very different
 - ▶ within variations: we observe a change in Y because there are changes in D within "units"
 - ▶ between variations: we observe a change in Y because there are changes in D between "units"

Practice Final Exam Questions II

Consider a regression $Y_{st} = \lambda_s + \theta_t + \rho D_{st} + X'_{st}\gamma + e_{st}$. What ID Assm do you have to make to interpret the OLS estimate of ρ as the cause effect?

► In words? In math/stat:

- $\{Y_{0st}, Y_{1st}\} \perp\!\!\!\perp D_{st} \mid \lambda_s, \theta_t, X_{st}$?
- $Y_{0st} \perp\!\!\!\perp D_{st} \mid \lambda_s, \theta_t, X_{st}$?
- $Y_{0st} - E(Y_{0st}) \perp\!\!\!\perp D_{st} \mid \lambda_s, \theta_t, X_{st}$?
- $E(e_{0st} \mid D_i, \lambda_s, \theta_t, X_{st}) = E(e_{0st} \mid \lambda_s, \theta_t, X_{st})$?
- $E(Y_{0it} - \bar{Y}_{0i} \mid D_i = 1, \theta_t, X_{st}) = E(Y_{0it} - \bar{Y}_{0i} \mid D_i = 0, \theta_t, X_{st})$?
- $E(e_{st} \mid D_i, \lambda_s, \theta_t, X_{st}) = 0$?
- $E(e_{st} D_i \mid \lambda_s, \theta_t, X_{st}) = 0$?
- $\text{Cov}(D_{st}^*, e_{st}) = 0$, where D_{st}^* is residuals from a regression of D_{st} on λ_s, θ_t and X_{st} ?

Practical Final Exam Question III

Write down a regression for a DDD type of design: age group 22-24 ($A_a = 0$) is not affected by the MLGA, while age group 18-22 ($A_a = 1$) is affected.

► How about

►
$$Y_{ast} = \lambda_s + \theta_t + \rho D_{st} + \beta_0 A_a + \beta_1 (A_a * \lambda_s) + \beta_2 (A_a * \theta_t) + \beta_3 (A_a * D_{st}) + e_{ast}?$$

►
$$Y_{ast} = \xi_{st} + \lambda_{as} + \theta_{at} + \pi(A_a * D_{st}) + e_{ast}$$

► How to think about complicated regression in general? Simplify variables as dummies and then write down the conditional expectations for different groups.

- In this case, assume two age groups, two states, and two years.
- For different age groups, you can perform a DID, what's left is a state-specific time effects

Follow-up Practical Questions

- ▶ How to run multiple-FE model in Stata or R:
 - ▶ If there is only one high-dimensional FE, absorb it using `areg`, `a(hi-group)` or `xtreg, fe` and then use LSDV, see Gormley and Matsa (2014);
 - ▶ `areg y x i.group#i.year, a(hi-group)`
 - ▶ `xtset hi-group; xtreg y x i.group#i.year, fe`
 - ▶ LSDV in R: `lm(y ~ d + factor(group))`
 - ▶ Linear Models for Panel Data in R: `library(plm)`
- ▶ If there are several high-dimensional FEs,
 - ▶ use `ssc install felsdsvreg` or `ssc install reghdfe` in Stata,
 - ▶ or use `install.packages("lfe")` in R

Practical Exam Question IV I

Interpreting multivariate OLS and quantile regression estimates (treatment group is full insurance, control group pays 20% of their medical expenses):

OLS	10 th	25 th	50 th	75 th	90 th
0.30 (0.04)	0.00 (0.05)	0.40 (0.04)	0.35 (0.04)	0.20 (0.04)	0.03 (0.05)

Practical Exam Question IV II

- ▶ The OLS estimate tells that, *conditional on X_i* , being in the treatment group raises the mean of log of medical expenses by 0.3, or about 30 percent.
- ▶ The quantile regression estimates indicate how various quantiles of the distribution of medical expenditures are affected by the treatment:
 - ▶ The 10th percentile of the treatment and control distribution are equal to one another, *conditional on X_i* .
 - ▶ The 25th percentile of the treatment distribution of log medical expenses is 0.4 larger than the 25th percentile of the distribution among the controls, conditional on X_i .
- ▶ Aside: Are these price elasticities?
 - ▶ How should we connect experimental estimates with theoretical construct?

Practical Final Exam Question V

What's the ID Asm for a FD/FE model:

$$Y_{it} = \alpha + \lambda_i + \beta X_{it} + e_{it},$$

or

$$Y_{i5} - Y_{i4} = \Delta Y_i = \beta \Delta X_i + \Delta e_i,$$

where Y_{i5} is test score in 5th grade, and X_{it} is teacher experience.

- ▶ Again?
- ▶ How about the lagged model:

$$Y_{it} = \alpha + \gamma Y_{i,t-1} + \lambda_i + \beta X_{it} + e_{it},$$

or the gain model:

$$Y_{it} - Y_{i,t-1} = \alpha + \lambda_i + \beta X_{it} + e_{it},$$

Practical Final Exam Question VI

Suppose that schools assign children who have unusually low test scores in one year to classroom in the following year that have more experienced teachers. Will this lead the OLS estimate of β in the model of ΔY_{it} to be upward biased, downward biased, or unbiased?

- ▶ Apply the OVB formula:

$$\frac{C(\Delta Y_{it}, \Delta X_{it})}{V(\Delta X_{it})} = \beta + \frac{C(\Delta e_{it}, \Delta X_{it})}{V(\Delta X_{it})}$$

- ▶ What's the sign of $C(\Delta e_{it}, \Delta X_{it})$?

- ▶ $C(\Delta e_{it}, \Delta X_{it}) = C(e_{i5} - e_{i4}, X_{i5} - X_{i4}) = C(e_{i5}, X_{i5}) - C(e_{i5}, X_{i4}) - C(e_{i4}, X_{i5}) + C(e_{i4}, X_{i4}) = -C(e_{i4}, X_{i5}) > 0$

Practical Final Exam Question VII I

Consider estimating the effect of school funding R_s on student achievement Y_{is} : $Y_{is} = \alpha + \rho R_s + X'_{is}\beta + Z'_s\gamma + e_{is}$. R_s is likely to be biased, why?

Suppose that beginning in 2010 a particular state gave money to local school districts using the following formula: If the fraction on students in 2009 who lived below the poverty level was greater than 35%, then the district would receive \$2000 per student per year. If the fraction was less than 35%, then the district would receive \$1000 per student per year. How would you use the new funding formula to estimate ρ ?

Practical Final Exam Question VII II

- ▶ Basically we have a fuzzy RD design. So
 - ▶ What's the running variable?
 - ▶ How would you define the dummy variable for the cutoff?
 - ▶ How would you model the relationship between the outcome variable and the running variable?
 - ▶ What's the ID Asm?
 - ▶ What're the implications of the ID Asm, or how can we test the validity of the Asm?
 - ▶ How to interpret the estimate of ρ ?
- ▶ Is it also possible to use a fuzzy DID design?

Reasons and Disclaimer

- ▶ Learning *latent* features of some texts is hard: many possible counterfactuals
- ▶ Perhaps it would be a good idea to develop a small routine/system/algorithm for reading
 - ▶ For those of you who plan to take the four applied micro fields, 75% of the second year assigned papers (5 to 15 every week) would be design-based; much like papers in Econ 535
 - ▶ Most of them are produced using “technologies” we learn this semester
- ▶ Disclaimer: I still don’t know how to read *effectively*
- ▶ Personal experience: the most effective learning strategy is doing research; teaching things you don’t know is also quite effective; taking class and reading passively are only as useful as watching TV

Things people do

- ▶ Answer a lists of question after reading
 - ▶ Example: See Robin Lee's [IO Syllabus](#)
- ▶ Compile literature notes: summary, intuition, discussion, etc.
 - ▶ [Example](#)
- ▶ Write haiku (for papers you read or write)
 - ▶ [Example](#)
- ▶ Organize data sources and variables
 - ▶ Example: DevEconData.blogspot.com
- ▶ Make notes while reading and then turn into your own sentences or paragraphs
 - ▶ Example: [Professor McCloskey's Little Lessons in Historigraphic Scholarship](#)

Split, Process, and Summarize

- ▶ Every paper contains a great amount of information
 - ▶ To extract these information in one round of reading is always daunting, at least for me
 - ▶ Sometimes it's easier to read one papers in many rounds, with different focus in each phrase
 - ▶ For every round of reading, process and summarize
- ▶ One advantage of this approach is that it makes the unit of information smaller, so that it can be substituted easily.
 - ▶ Ref

Beware of Initiative Bias

- ▶ According to Glaeser (2006), researcher's problem is

$$\max_E \theta S(f(E) + R) - E,$$

where θ is the exogenous importance of the research question, $S(\cdot)$ reflects the extent the public attributes importance to the result, R is the real significance of the research, which is a r.v. follows some distribution $g(R)$, E is the non-negative effort of the researcher, and $f(\cdot)$ is the production function

- ▶ The FOC is $\theta * S'(f(E) + R)f'(E) = 1$.
- ▶ There are many interesting predictions from his model, but his first point is that *“we should accept researcher initiative as being the norm, and not the exception”*.

Some Suggestions I Heard and I Wish I can Always Follow

- ▶ Compile a list of questions and force yourself to write down answers in full sentences after reading a paper
 - ▶ For example, what's the source of variations they use; what's the ID Asm; what exactly is the comparison groups; is their regression the one they want; what restrictions does their regression imposes. . .
- ▶ Pick two papers every week and read it word-by-word, figure out everything (narratives, equations, summary statistics, estimates) you're not comfortable with
 - ▶ Ask your classmates, TAs, and Professors
- ▶ My biggest lesson from this year's TA experience: I am constantly surprised by my own ignorance.

References

Glaeser, Edward L. 2006. *Researcher Incentives and Empirical Methods*. Working Paper 329. Technical Working Paper Series. National Bureau of Economic Research. doi:[10.3386/t0329](https://doi.org/10.3386/t0329).

Gormley, Todd A, and David A Matsa. 2014. "Common Errors: How to (and Not to) Control for Unobserved Heterogeneity." *Review of Financial Studies* 27 (2). Soc Financial Studies: 617–61.