ECON 251 Office Hours

Review + Extensions

Elird Haxhiu

Fall 2022

Course Evaluations

- Please do them :)

Outline

1. Review: HW3 solutions + discuss HW4
2. Competing Paradigms: Reductionism vs Chaos/Complexity
3. Review: OLS, IV, DID, LPM

4. Sensitivity analysis (Cinelli & Hazlett, 2020)
5. Directed Acyclic Graphs (DAGs) for intuition, and identification
6. Bayesian methods in econometrics
7. Recommendations for further reading

Review HW3 solutions

Discuss HW4

- Due tonight: Thursday December 8th @ 11:59pm

Extra Problems

- Posted tomorrow
- We will do them together next week
- A “practice exam” let’s say ;)

Outline

1. Review: HW3 solutions + discuss HW4
2. Competing Paradigms: Reductionism vs Chaos/Complexity
3. Review: OLS, IV, DID, LPM

4. Sensitivity analysis (Cinelli & Hazlett, 2020)
5. Directed Acyclic Graphs (DAGs) for intuition, and identification
6. Bayesian methods in econometrics
7. Recommendations for further reading

Chaos/Complexity vs Reductionism

- How do we interpret the error term U in our linear model of the outcome

$$Y = \beta_0 + \beta_1 X + U$$

1. Deviation from the system, error that is ignorable due to crude instrument devices or crude theory… measurement error
2. An inherent part of the system we must study and understand, more variability does not mean more ignorance

- EX: measuring temperature
- EX (not just about measurement): does testosterone make aggressive?

Sapolsky (2011) lecture

https://www.youtube.com/watch?v=_njf8jwEGRo

- Under reductionism, a more refined theory or more precise measurement should always reduce variance (increase precision)
- Key tools of reductionism = linearity + additivity
- Chaos = complex systems exhibit non-linearities & non-additivities
- Implies "butterfly effect" = small changes in one part of the system can amplify to have large effects on overall state (equilibrium)
- Periodicity vs aperiodicity = both can arise from deterministic rules, giving rise to either reductive tools (since time linear, can forecast any future state given initial conditions) or chaotic tools (impossible to forecast, must go step-by-step)
- Mincer (1974) model that gives rise to $\log Y = \alpha + \beta \cdot S + U$ is reductive model

Sapolsky and Bolt (1996)

- Does variability decrease with more reductive approaches?
- Conduct meta-study of (all?) papers estimating effect of testosterone on aggressive behavior

- Four measures drawn from each paper (since 1987)
    1. % figures with quantitative data
    2. % of quant data shown with standard errors (SE)
    3. Mean coefficient of variation (CV) for all data (inverse of t-stat!)
    4. # citations

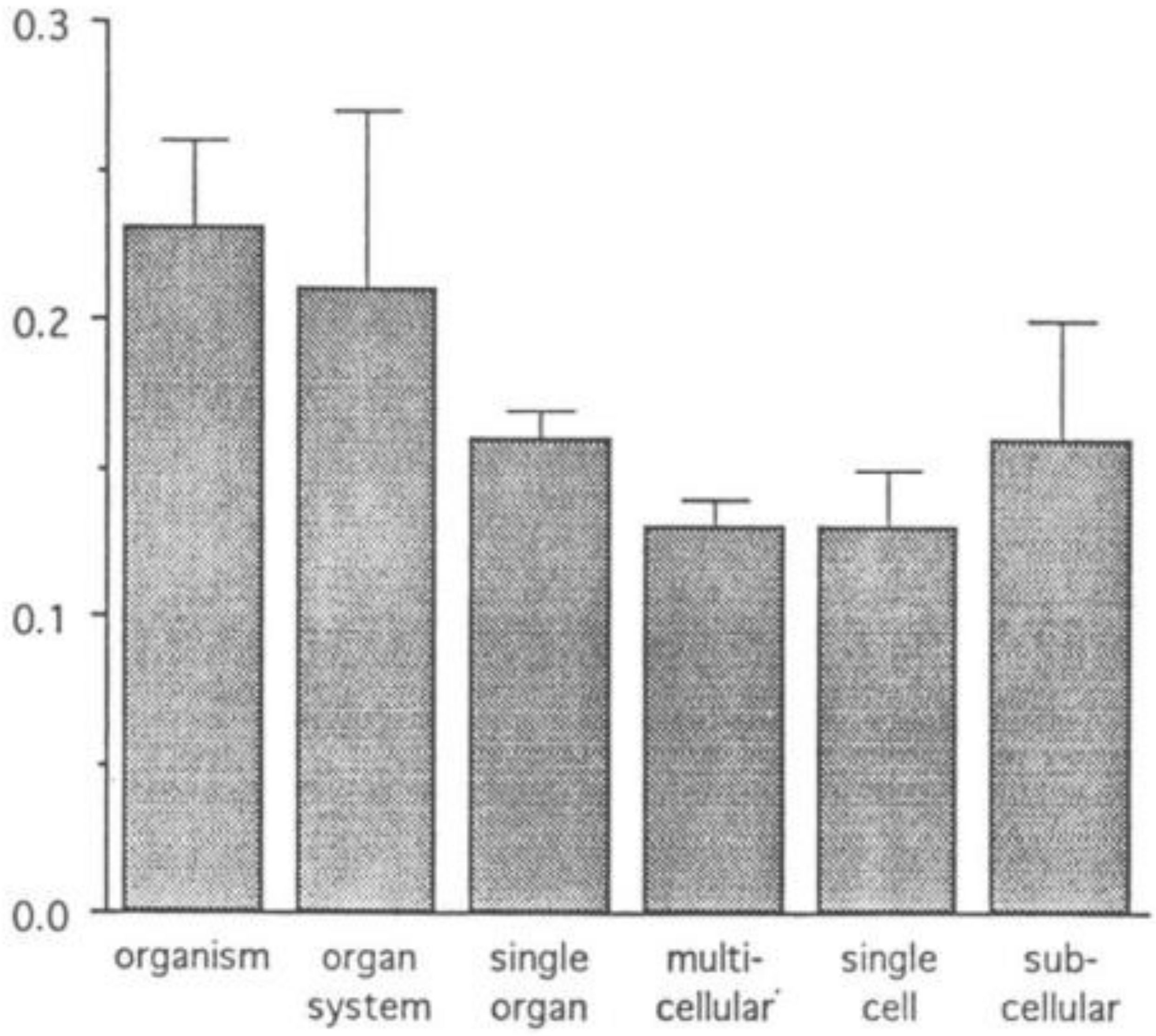


FIG. 3.—Average coefficient of variations of all papers, as a function of reductive level of techniques.

Criticisms?

1. Meta-analysis usually within, not across, disciplines… comparing “apples and oranges” but… they document similarities between them!
2. More fundamental (imo): there may be different types of confounders to the relationship at different levels of aggregation…

Review

1. OLS assumptions and theorems
2. IV assumptions and results
3. DID assumptions and results
4. LPM intuition and results

Usual Assumptions

- MLR1 (linear outcome model) $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$
- MLR2 (random sampling) $\{Y_i, X_{i1}, \ldots, X_{ik}\}_{i=1}^{N}$ is random draw
- MLR3 (no collinearity) no X_{ij} linear function of any other X_{il}

- MLR4 (independence) $E[U_i | X_{i1}, \ldots, X_{ik}] = 0$

- MLR5 (homoskedasticity) $\mathrm{Var}(U_i | X_{i1}, \ldots, X_{ik}) = \sigma^2$
- MLR6 (normality) $U_i \sim N(0, \sigma^2)$

$$\Rightarrow Y_i \sim N(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}, \sigma^2)$$

Ordinary Least Squares (OLS) Estimator + Results

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_k\}} \frac{1}{N} \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2 \qquad \Rightarrow \qquad \hat{\beta}_j^{OLS}$$

- T1 (unbiased) MLR1+2+3+4 $\Rightarrow$ $E\left[\widehat{\beta}_j^{\,OLS}\right] = \beta_j \qquad \forall j = \{0,1,\ldots,k\}$
- T2 (efficient) MLR1+2+3+4+5 $\Rightarrow$ $E\left[\widehat{\beta}_j^{\,OLS}\right] = \beta_j \qquad \forall j = \{0,1,\ldots,k\}$

  (Gauss-Markov) $\mathrm{Var}\left[\widehat{\beta}_j^{\,OLS}\right] \leq \mathrm{Var}\left[\widehat{\beta}_j^{\,\text{other linear}}\right]$

Ordinary Least Squares (OLS) Estimator + Results

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_k\}} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2 \qquad \Rightarrow \qquad \hat{\beta}_j^{OLS}$$

- T3 (efficient) MLR1+2+3+4+5+6 $\Rightarrow$ $\widehat{\beta_j}^{OLS} \sim N\left(\beta_j, \mathrm{Var}\left[\beta_j\right]\right) \quad \forall j = \{0,1,\ldots,k\}$

  (Classical)

$$\frac{\widehat{\beta_j}^{OLS} - \beta_j}{\mathrm{sd}[\beta_j]} \sim N(0,1) \qquad \frac{\widehat{\beta_j}^{OLS} - \beta_j}{\mathrm{se}[\beta_j]} \sim t(N-k-1)$$

Omitted variable bias (OVB)

“True” model $\log Y_i = \alpha + \beta \cdot S_i + \delta^{X \to Y} \cdot X_i + U_i$ $\quad \mathrm{Cov}(S_i, U_i) = 0$

Our model $\log Y_i = a + b \cdot S_i + E_i$

Auxiliary model $X_i = c + \gamma^{S \to X} \cdot S_i + \eta_i$

In week 5 we proved that naively assuming $\mathrm{Cov}(S_i, E_i) = 0$ in our model implies

$$b = \frac{\mathrm{Cov}(S_i, \log Y_i)}{\mathrm{Var}(S_i)}$$

$$= \beta + \gamma^{S \to X} \cdot \delta^{X \to Y}$$

$=$ causal effect + (var in S related to X) $\cdot$ (var in X related to Y)

Derive MM estimator under IV assumptions

Start w/ IV exogeneity $\text{Cov}(Z_i, U_i) = 0$ + substitute Mincer (1974) earnings model

$$\text{Cov}(Z_i, \log Y_i - \alpha - \beta \cdot S_i) = 0$$
$$\text{Cov}(Z_i, \log Y_i) - \text{Cov}(Z_i, \alpha) - \beta \cdot \text{Cov}(Z_i, S_i) = 0$$
$$\text{Cov}(Z_i, \log Y_i) = \beta \cdot \text{Cov}(Z_i, S_i)$$
$$\beta = \frac{\text{Cov}(Z_i, \log Y_i)}{\text{Cov}(Z_i, S_i)}$$

$$\Rightarrow \hat{\beta}^{MM} = \frac{\widehat{\text{Cov}}(Z_i, \log Y_i)}{\widehat{\text{Cov}}(Z_i, S_i)} := \frac{\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)\left(\log Y_i - \overline{\log Y}\right)}{\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)\left(S_i - \overline{S}\right)}$$

Consistency of MM estimator under IV assumptions

- Start from definition of estimator, and compute the probability limit

$$\hat{\beta}^{MM} = \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)} = \frac{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$= \frac{\mathrm{Cov}(Z_i, \log Y_i)}{\mathrm{Cov}(Z_i, S_i)} = \frac{\mathrm{Cov}(Z_i, \alpha + \beta \cdot S_i + U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

$$= \frac{\mathrm{Cov}(Z_i, \alpha) + \beta \cdot \mathrm{Cov}(Z_i, S_i) + \mathrm{Cov}(Z_i, U_i)}{\mathrm{Cov}(Z_i, S_i)} = \beta + \frac{\mathrm{Cov}(Z_i, U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

Doing IV can be worse than OLS

- The OVB formula for OLS implies that it converges to

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{OLS} = \beta + \frac{\mathrm{Cov}(S_i, U_i)}{\mathrm{Var}(S_i)}$$

- We have just shown that the MM estimator converges to

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \beta + \frac{\mathrm{Cov}(Z_i, U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

- What if $\mathrm{Cov}(S_i, U_i) = 0$ or $\mathrm{Cov}(Z_i, U_i) = 0$ are not exactly $= 0$?
- Not clear which is more likely to hold without more context, but...
- If $\mathrm{Cov}(Z_i, S_i) \approx 0$ (weak instrument), then even minor violations of IV exogeneity lead to large asymptotic bias: aka inconsistency!

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

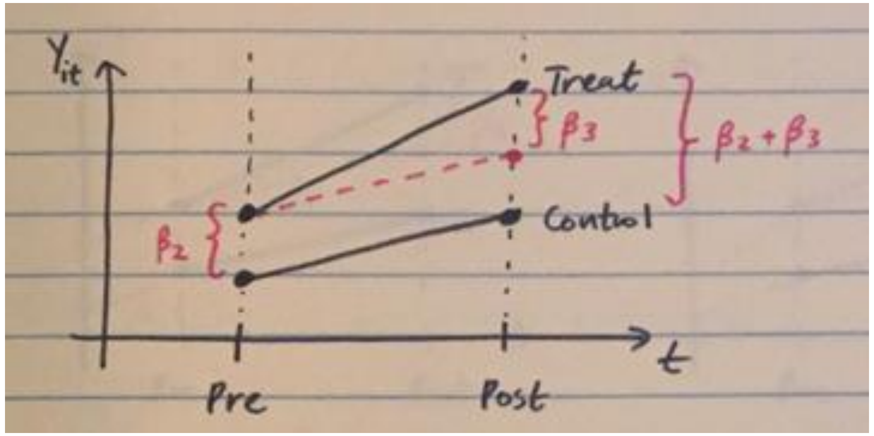
2 periods (before/after) and 2 groups (treated/control)

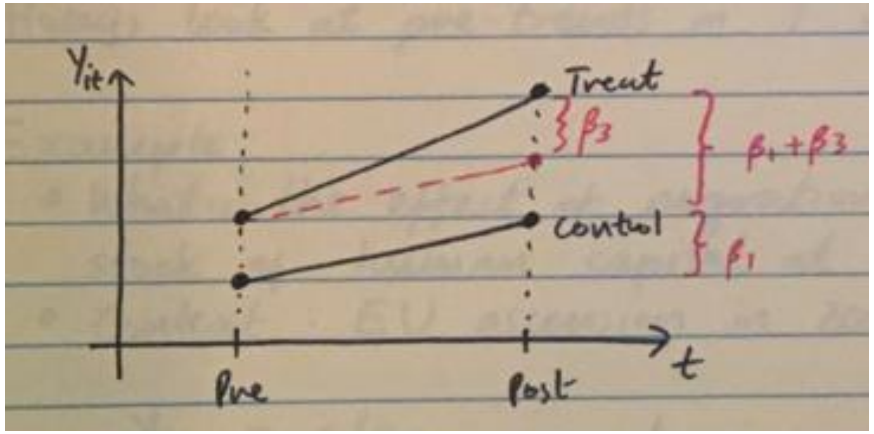
$Y_{it} :=$ outcome of interest

$P_t := 1\{t \text{ is after treatment occurs}\}$

$T_i := 1\{i \text{ is treated/exposed}\}$

$$Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 [P_t \cdot T_i] + U_i$$

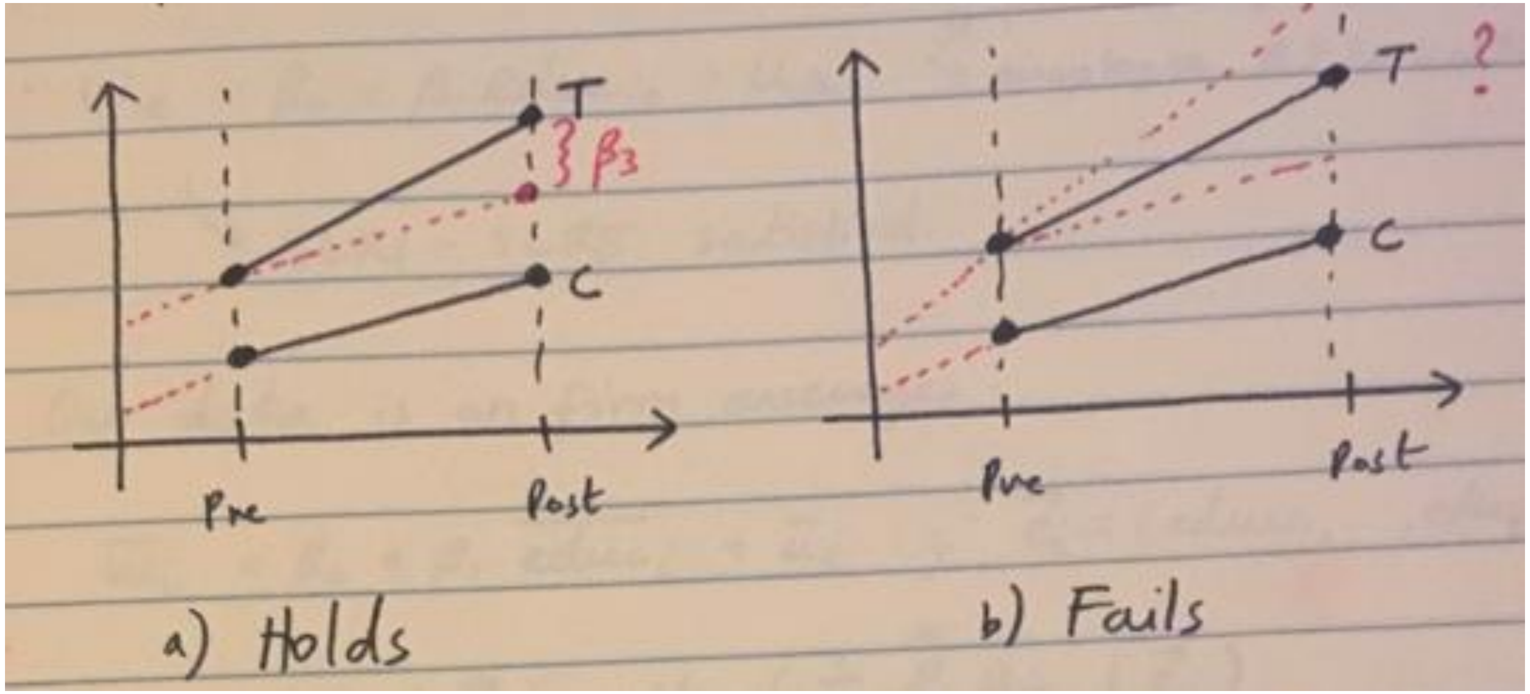




	Before	After	After – Before
Control	β_0	$\beta_0 + \beta_1$	β_1
Treated	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Treat – Control	β_2	$\beta_2 + \beta_3$	β_3

Parallel Trends Assumption = exposed units Y without policy T would have changed like unexposed units Y

- PTA is an untestable assumption, just like OLS exogeneity or IV exogeneity
- However, if we have access to more data before policy, we can assess how likely it is to hold in practice… commonly known as "checking for pre-trends"
- One reason why people seem to like DD… visual check of identifying assumption!



Linear Probability Model (LPM) - Binary Outcomes $Y \in \{0,1\}$

- Independence + binary Y gives "change in prob(Y=1)" interpretation

$$Y = \beta_0 + \beta_1 X + U$$

$$\begin{aligned} E[Y|X] &= P[Y = 1|X] \cdot 1 + P[Y = 1|X] \cdot 0 \\ &= P[Y = 1|X] \end{aligned}$$

$$\Rightarrow \beta_1 = \frac{\partial}{\partial X} P[Y = 1|X]$$

- OLS estimates of linear model with binary outcome $\Rightarrow$ LPM

Linear Probability Model (LPM) - Binary Outcomes
$Y \in \{0,1\}$

- LPM is nice because…
    1. Easy to estimate
    2. Easy to interpret

- LPM is problematic because…
    1. Predicted values of outcome can be outside of $[0,1]$ interval
    2. Does not make sense for X to change $P[Y = 1|X]$ linearly
    3. Homoskedasticity is always violated

$$
\begin{aligned}
\mathrm{Var}(Y|X) &= E[Y^2|X] - E(Y|X)^2 \\
&= [P(Y = 1|X) \cdot 1^2 + P(Y = 0|X) \cdot 0^2] - [P(Y = 1|X)]^2 \\
&= P(Y = 1|X) - P(Y = 1|X)^2 \\
&= P(Y = 1|X)[1 - P(Y = 1|X)]
\end{aligned}
$$