Lecture 1 – Introduction

Economics 8379
George Washington University

Instructor: Prof. Ben Williams

Let θ denote a parameter of interest.

- estimator, $\hat{\theta}_n$
- standard error, $SE(\hat{\theta}_n)$
- confidence interval, $\hat{\mathcal{C}}_n^\alpha$

Inconsistency (large sample bias): $\hat{\theta}_n \nrightarrow_p \theta$

Inconsistency (large sample bias): $\hat{\theta}_n \nrightarrow_p \theta$

- examples:
  - misspecified model
  - endogeneity
  - OLS estimator if X_i is correlated with u_i
- solutions?
  - a correctly specified model!
  - robust estimator
  - estimation/inference using bounds/identified set approach

Consistent but biased: $\hat{\theta}_n \to_p \theta$ but $E(\hat{\theta}_n) \neq \theta$

Consistent but biased: $\hat{\theta}_n \rightarrow_p \theta$ but $E(\hat{\theta}_n) \neq \theta$

- examples:
    - OLS estimator for an AR model
    - IV estimator under standard assumptions
    - MLE
- solutions?
    - not necessary if n is sufficiently large...
    - solutions based on better asymptotic approximations

Unbiased but standard errors are inconsistent:
$Var(\hat{\theta}_n)^{1/2}/SE(\hat{\theta}_n) \nrightarrow_p 1$

Unbiased but standard errors are inconsistent:
$Var(\hat{\theta}_n)^{1/2}/SE(\hat{\theta}_n) \nrightarrow_p 1$

- examples:
  - heteroskedasticity in errors, conventional standard errors used
  - autocorrelation in errors, heteroskedastic-robust standard errors used
- solutions:
  - HAC standard error formula!

Unbiased, standard errors are consistent, but $Pr(\theta \in \hat{C}_n^{\alpha}) \neq 1 - \alpha$

- examples:
  - variance estimate is biased though consistent
  - small sample problem – exacerbated by fat tails
- solutions?
  - bias corrections to standard error formulas
  - use more conservative critical values
  - bootstrap

Bootstrap

- Primer on the bootstrap
  - the idea is to use the empirical distribution rather than the asymptotic distribution
  - two main advantages
    - asymptotic refinements possible
    - easier when an analytical formula for standard errors is difficult
  - warning: in some econometric models (nonsmooth models) the bootstrap is not consistent

Bootstrap

- Bootstrap distribution:
  - randomly sample from the n observations n times (with replacement)
  - compute the statistic on this bootstrap sample
  - repeat this B times
  - approximate the distribution of the statistic using B observations of it
- asymptotic refinements can occur when the statistic is *pivotal*
- for regression: resample (y_i, X_i) or just the residual

Bootstrap

- in the case of heteroskedasticity:
  - residual bootstrap is not valid; pairs bootstrap is but does not provide refinement
  - the wild bootstrap:
    - resample: $y_i^* = X_i'\hat{\beta} + HC_s\hat{e}_i\nu_i^*$ where ν_i^* is, for example 1 or -1 with equal probability.
- see MacKinnon's notes on wild bootstrap for more.

Bootstrap

- clustering, stratification, etc.:
  - in a simple clustering setup: block bootstrap
    - only sample clusters/blocks
  - more generally, mimic the sample design in the bootstrap sampling
  - typically does not provide asymptotic refinement but avoids complicated standard error derivations
- more: parametric bootstrap, recentering and rescaling, testing in overidentified models
- see Chapter 11 in CT, 2001 Handbook Chapter by Horowitz, and brief discussions in Deaton (1997) and AP

Bootstrap - one more issue

- the bootstrap provides a refinement when tails are thin enough
- in some cases with heavy-tailed, asymmetric distribution, the bootstrap does just as poorly as the asy. approximation
- *not* an issue of biased variance estimate
- Bahadur-Savage (1956) impossibility.

Consistent but (asymptotically) inefficient:
$lim_{n\to} Var(\hat{\theta}_n) > lim_{n\to} Var(\tilde{\theta}_n)$ for some $\tilde{\theta}_n \to_p \theta$.

- examples:
    - OLS under heteroskedasticity
- solutions?
    - find the efficient estimator (WLS)
    - sacrifice efficiency to avoid misspecification bias

Model selection (pre-testing) distorts inference:
$Pr(\theta \in \hat{\mathcal{C}}_n^\alpha) \approx 1 - \alpha$ but $Pr(\theta \in \hat{\mathcal{C}}_n^\alpha \mid \hat{\mathcal{M}}) \not\approx 1 - \alpha$

- examples
  - regression specification with many regressors
  - pre-trend test in diff-in-diff estimator
- solutions?
  - active literature in econometrics

Econometric causality

- Chapters 1 and 2 in MHE and Heckman (2008) both address this issue.
- There is probably more agreement than disagreement between these two readings.
- I will stick primarily to Heckman (2008) today.

Three tasks

- Heckman (2008) argues for separating three tasks involved in causal analysis:

Table 1
Three Distinct Tasks Arising in the Analysis of Causal Models

Task	Description	Requirements
1	Defining the set of hypotheticals or counterfactuals	A scientific theory
2	Identifying causal parameters from hypothetical population data	Mathematical analysis of point or set identification
3	Identifying parameters from real data	Estimation and testing theory

MHE FAQs

1. What is the causal relationship of interest?
2. the experiment that could ideally be used to capture the causal effect of interest
   - “research questions that cannot be answered by any experiment are FUQs: fundamentally unidentified questions”
3. What is your identification strategy?
4. What is your mode of statistical inference?

- Heckman (2008) argues that: “Many ‘causal models’ in statistics are incomplete guides to interpreting data or for suggesting answers to particular policy questions. They are motivated by the experiment as an ideal. They do not clearly specify the mechanisms determining how hypothetical counterfactuals are realized or how hypothetical interventions are implemented except to compare ‘randomized’ with ‘nonrandomized’ interventions. They focus only on outcomes, leaving the model for selecting outcomes only implicitly specified. The construction of counterfactual outcomes is based on appeals to intuition and not on formal models.”

- He emphasizes the provisional nature of causal knowledge, given that the models required to define the causal effect are provisional.
- Does this conflict with Angrist and Pishke's view?

Types of policy problems

- Three policy evaluation problems:
  P1. "Evaluating the impact of historical treatments on outcomes..."
  P2. "Forecasting ... the impacts of interventions implemented in one environment in other environments..."
  P3. "Forecasting the impacts of interventions ... never historically experienced to various environments..."

Types of policy problems

- Structural models hold out the hope of answering P1-P3.
- “Reduced form” models can only answer P1.
- “By focusing on one narrow black box question, the treatment effect literature avoids many of the problems confronted in the econometrics literature that builds explicit models of counterfactuals and assignment mechanisms. This is its great virtue. At the same time, it produces parameters that are more limited in application.”

Potential outcome framework

- two counterfactual outcomes: Y_{0i} and Y_{1i}
- let D_i indicate treatment status of individual i
- only observe $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ for a random sample
  - there is an invariance assumption implicit here: $Y_i = Y_{iD_i}$
  - in statistical literature this is called SUTVA
- selection bias: $E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0) =$ $E(Y_{1i} - Y_{0i} \mid D_i = 1) + E(Y_{0i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0)$
- if D_i is randomly assigned?

The evaluation problem

- The individual level treatment effect: $Y_{1i} - Y_{0i}$
- When we write this down, we implicitly assume a type of policy invariance – the potential outcomes don't depend on the treatment assignment mechanism (see pages 6-8 in Heckman, 2008).
- The evaluation problem: we never observe $Y_{1i} - Y_{0i}$ for any i.
- Two solutions to the evaluation problem:
  - The "structural" approach: model the determinants of Y_{1i}, Y_{0i}, D_i, including any dependence between Y_{di} and D_i.
  - The "treatment effect" approach: ignore determinants of outcomes and focus on estimating means of $Y_{1i} - Y_{0i}$

- An example of the structural approach:

$$Y_1 = X\beta_1 + U_1$$
$$Y_0 = X\beta_0 + U_0$$
$$C = Z\gamma + U_C$$

  and $D = \mathbf{1}(E(Y_1 - Y_0 - C \mid \mathcal{I}) \geq 0)$ where $\mathcal{I}$ is the individual's information set.
- Within this model we can answer a lot of interesting economic questions.
  - E.g., we can distinguish between ex ante and ex post treatment effects.
- Whether we can identify answers to these causal/policy questions given a particular source of data is a separate question.

Marschak's maxim

- Marschak's maxim: "formulate the problem being addressed clearly and ... use the minimal ingredients required to solve it."
- The "treatment effect" approach is a particular application of this maxim.
  - "For certain classes of policy interventions, designed to answer problem P1, the treatment effect approach may be very powerful and more convincing than explicitly formulated models because it entails fewer assumptions."
- But it answers a fairly limited set of policy questions and often the particular policy questions being answered – and why it is important – is not addressed by the "treatment effect" approach.

Conclusion

- Heckman (2008) concludes by saying that as the structural approach provides new methods of identification that relax strong assumptions and the treatment effect literature expands the set of policy counterfactuals it seeks to evaluate, the two approaches will merge.
- Heckman (2010) expands on this.
- This will be an overarching them of this class as we visit both “structural” and “treatment effect” or “reduced form” methods.

Conclusion

The course will be organized broadly according to two characteristics of models/estimators:

- relaxing assumptions regarding structure (nonlinearity, heterogeneity, simultaneity, etc.)
- various assumptions to address endogeneity (unconfoundedness, IV, RD, panel FE/DD, etc.)