Discussion of Pre-Specified Analysis Plans by Peterson et al.

Bryan Graham

University of California - Berkeley and NBER

07 December 2012
Some Basic Elements

• Let $Y$ be an outcome, $X$ a policy variable and $W$ additional unit-specific covariates

• A population is a joint distribution $F_{W,X,Y}(w,x,y)$

• An estimand is some function of this joint distribution.

  – e.g., a difference in partial means

  $\beta(x',x) = \mathbb{E}_W[\mathbb{E}[Y|W,X = x'] - \mathbb{E}[Y|W,X = x]]$
Some Basic Elements (continued)

• **Reification** of estimand involves argument (e.g., randomization, ‘as if’ randomization)

• If \( x' = 1 \) and \( x = 0 \), randomly assigned, \( \beta(x', x) \) is an Average Treatment Effect (ATE)

• **Estimation/inference** requires a statistical model for \( F_{W,X,Y}(w, x, y) \)
Some Comments

• All steps important, last two especially tricky and fraught with pitfalls

• **Problem:** If $W$ is high dimensional a non-parametric treatment of $F_{W,X,Y}$ is impractical; semiparametric modelling, as it requires the a priori imposition of restrictions on the DGP, involves ‘art’.

• In ATE example the statistical/inference issues is one of covariate adjustment with a potentially high dimensional conditioning vector (this is a well-posed research topic)

• Reification step is typically heroic in observational work
  
  – ...but estimand may still be interesting (apples-to-apples comparisons)
Final Comments

- We typically deal with broken randomized experiments (attrition, refusal etc.) via covariate adjustment
  - Under a missing-at-random (MAR) type assumption this adjustment is a statistical matter

- Partial identification may be a useful complement to invoking MAR in this setting.
Final Comments (continued)

- Research will be easier to evaluate if above steps are clearly separate.
  - population – sampling frame/design
  - estimand – functional of a joint distribution
  - reification – involves research design and/or persuasion
  - estimation/inference – well-defined problem, but hard!