Putting the Evidence in Evidence-Based Policy

Jeffrey Smith
University of Michigan

State of the Art Lecture
Canadian Economic Association Meetings
Ottawa, Ontario
June 2011
Gratitude

Arthur Sweetman in particular

Canadian economics more generally
Outline

Evaluation policies
Parameters of interest
Types of evidence / rankings and comparisons
General equilibrium effects
Cost-benefit analysis
Data quality
Evaluation policy
Alternatives to econometric evaluation
Institutions
Conclusions
Potential outcomes framework

\[ Y_{0i} = \text{untreated outcome for unit "}i\text{"} \]

\[ Y_{1i} = \text{treated outcome for unit "}i\text{"} \]

\[ Y_{1i} - Y_{0i} = \Delta_i = \text{impact for unit "}i\text{"} \]

Note that the impact varies across units!

This is intuitive in most cases but often ignored in studies that focus on “the effect” of some policy.
Parameters of interest

Old standbys
  Treatment on the treated: \( E(Y_{1i} - Y_{0i} \mid D = 1) \)
  Average treatment effect: \( E(Y_{1i} - Y_{0i}) \)

New kids on the block
  Local average treatment effect (more later)
  Quantile treatment effects
  Variance of impacts

Different parameters correspond to different policy questions.

See e.g. Djebbari and Smith (2008) *Journal of Econometrics*
Heterogeneous treatment effects

Treatment effect may also vary with

Features of the treated units
Features of the policy environment
Features of the economic environment

Example: U.S. Job Training Partnership Act and barriers to work

Example: active labor market policy over the cycle, as in Lechner and Wunsch (2009) *Journal of Labor Economics*

This variation is important for understanding how treatments work and for targeting (as in SOMS) either formally or informally
Experiments

Random assignment makes causal claims very compelling indeed!

Example: IES intensive teacher mentoring experiment

Example: IES abstinence-only sex education experiment

Experiments are easy to explain to politicians

Issues with experimental evaluations

Experiments can be conducted well or poorly and can sometimes be manipulated by program staff

Substitution and dropout change the parameter being estimated and complicate interpretation

Experiments provide only partial equilibrium estimates
Recent developments in experiments

Econometric advances with group level random assignment

Using experiments to test or calibrate structural models

   Ex: Todd and Wolpin (2005) and Lise et al. (2011)

Using experiments to estimate structural parameters

   Back to the future example: US NIT experiments

   Ex: Ludwig et al. (2011)
The bottom lines on experiments

Burt Barnow: experiments are not a substitute for thinking!

But do more experiments!

Or, in the case of Canada: do some experiments!
Selection on observed variables

Random assignment conditional on observed variables!

Is the goddess this kind to evaluation researchers?

Important implications for the choice of evaluation policy and, by extension, for data collection
Key conceptual issue: what observed variables are required?

Depends on policy type and on institutions

Experiments are helpful here, as is thinking about the economics and reading the literature

Defines an ongoing research strategy of collecting new types of variables suggested by theory and seeing if they matter
  Ex: WIA / LEHD project by Holzer, et al.
  Ex: Lechner and Wunsch (2011)

General lesson: lagged outcomes really important
  Ex: Dolton and Smith (2011)
Matching, weighting and regression

Parametric linear regression
   Angrist (1998): OLS does not estimate ATET!

Nearest neighbor (or pair) matching
   Rules the statistics literature
   Optimal full matching as in e.g. Hansen (2004)

Kernel matching and related schemes
   Estimating a non-parametric regression of $Y_0$ on $P(X)$
   Can also frame was a weighting estimator
Matching, weighting and regression (continued)

Inverse propensity weighting
   Horvitz and Thompson (1952) *JASA*
   Attains the semi-parametric efficiency bound
   Watch out for probabilities near zero and one
   Inverse probability tilting (odd name, nice estimator)

Important recent Monte Carlo analyses
   Busso, DiNardo and McCrary (2009a,b)
   Huber, Lechner and Wunsch (2011)

Justin McCrary: matching is an attempt to approximate what reweighting is doing directly
Dynamic treatment assignment

Going beyond Heckman and Robb (1985): What if treatment is offered in more than one period?

Paradigmatic example: active labor market programs in Europe

No-treatment comparison group implicitly conditions on future outcomes

Solution (?): estimate the impact of treatment now versus waiting, where waiting may imply future treatment

Instrumental variables

“Looking for instruments”

Sometimes instruments can be designed into the institutions

Random assignment is the ideal instrument

RD and D-in-D can be framed as IV estimators

Key issue 1: instrument validity

False Folk Theorem of Empirical Economics #9

Ex: Buckles and Hungeman (2010) on quarter of birth
Instrumental variables (continued)

Key Issue 2: instruments identify Local Average Treatment Effects

When is the LATE interesting?

Example: instruments based on the cost of participation
Example: instruments based on multiple births
Example: mandatory schooling laws
Example: Conley and Dupor (2011)

A bad instrument may be worse than no instrument

See the recent exchanges between Imbens, Heckman and Deaton
Longitudinal methods

Includes “diff-in-diff” and more general panel methods

Relies on “bias stability” assumption, wherein selection into treatment depends only on time-invariant unobserved differences across units

Can be “designed in” via staged rollout of treatments over multiple administrative units
  Ex: Progresa evaluation in Mexico
  Ex: Healthy Babies, Healthy Children in Ontario

But, does Canada need more provinces?
Longitudinal methods (continued)

Issues with current applied econometric practice:

Treatment timing depends on transitory shocks to outcomes
    Example: “Ashenfelter’s Dip” for ALMP participants

Changes in treatment effects over time relative to policy initiation
    Example: Card and Krueger minimum wage study
    Example: tax reform difference-in-differences studies

Anticipatory effects of treatment

Regression discontinuity

Design RD into the institutions – must be enforced!

Need population at the RD

Only identifies treatment effect at the discontinuity

Beware of age and time discontinuities (think about economics)

Example: Reading First (IES)
Example: Healthy Babies, Healthy Children in Ontario
Methods advances in RD

Power calculations for RD designs
  Schochet (2008) IES report

Bandwidth selection
  Imbens and Kalyanaraman (2009)

Specification testing

Discrete running variable
  Lee and Card (2007) *Journal of Econometrics*
RD surveys


Lee and Lemieux (2010) *Journal of Economic Literature*
Comparisons of econometric evaluation methods

Much of the literature asks the wrong question.

What we should want: mapping from parameter, data, and institutions to choice of econometric method (or stop)

This mapping is not degenerate! There is no magic bullet.

Example: NSW literature, e.g. LaLonde (1986), Heckman and Hotz (1989), Dehejia and Wahba (1999, 2002), Smith and Todd (2005a,b)

Not about “does matching work” but about whether the NSW data suffice for identification with various econometric estimators
Rankings of econometric evaluation methods

US IES What Works Clearinghouse:

Three grades (plus omission from the clearinghouse)
Only experiments and RD papers can get the top grade
Thoughtful, clearly-defined written evaluation criteria
Multiple evaluators for every study + resolution
Formal appeals process

Lots of “within” variation as well as “between” variation

How much of the overall variation in quality is explained by identification strategy is an interesting empirical question
General equilibrium effects

SUTVA and partial equilibrium analysis

Implications for data collection

Identification strategy 1: Geographically dispersed markets
  Ex: Forslund and Krueger (1992)
  Ex: Angelucci and di Giorgio (2008)

Identification strategy 2: Structural models
  Example: Davidson and Woodbury (1992)
  Example: Heckman, Lochner and Taber (1998)
  Example: Lise, Seitz and Smith (2009)
Cost-benefit analysis

Discount rates

Excess burden

How long do the impacts last?
  Example: National Supported Work Demonstration
  Example: National Job Corps Study
  No easy answer – be nice to have more data points

Incorporate all relevant outcomes
  Example: crime in the National Job Corps Study
  Example: fertility in Germany; Lechner and Wiehler (2007)
  Some may be difficult to monetize
Cost-benefit analysis (continued)

Need good data on costs
  Canada could do (much) better here
  So could everyone else
  Ex: Farrell’s Ice Cream Parlour Restaurant

Incorporate general equilibrium effects
  Use the literature as a guide if no study-specific estimates

Heckman, LaLonde and Smith (1999)

Some programs do not work – the “Nothing Works Clearinghouse”
Evaluation policy

Three broad approaches and their issues:

1. Do experiments
   Expensive
   Politically challenging
   Not everything can be randomized

2. Rich administrative data (think Denmark or Germany)
   Strategic reporting (IPEDS in the US)
   Empty fields (education in Canada)
   Imperfect linking
   May still need to supplement with surveys
   Privacy issues
Evaluation policy (continued)

3. Survey based evaluations
   Can link surveys to administrative data
   Response rates falling on voluntary surveys
   Political costs of mandatory surveys
   Sampling getting harder over time
   Response mode issues
   Item non-response

Canada is implicitly pursuing the third strategy.

A blunt discussion of the issue at this level would be valuable.
Alternatives to econometric program evaluation

Charlatans

Performance management

Heinrich, Heckman and Smith (2002), Heckman et al. (2011)

Customer satisfaction / participant self-evaluations

Example: Teaching American History
Example: Smith, Whalley and Wilcox (2011)
Institutions

Public use (or research use) data sets

Peer review – HRSDC doing well here

Incentives for consultants to publish in academic journals

Seconding academics to government agencies (e.g. chief economist at US DOL)

Research conferences for academics, policy people and consultants

Independent (more or less) evaluation shops (e.g. IFAU)
Conclusions

The truth is out there!

There is no such thing as “the effect” of a policy

There is no substitute for thinking about identification

Worry about general equilibrium effects

Get the cost/benefit analysis right; avoid “one number”

No easy alternatives to econometric policy evaluation

Institutions matter