MICROECONOMETRIC EVALUATION WITH OBSERVATIONAL DATA:

AN **INCOMPLETE** SURVEY

Michael Lechner

Swiss Institute for International Economics and Applied Economic Research (SIAW)
University of St. Gallen

Halle, December 1999

- Introduction
- Notation and definition of effects
- Selection bias and different identification strategies
- Conclusions

Email: Michael.Lechner@unisg.ch

WWW: http://www.siaw.unisg.ch/lechner

"Social scientists never have access to true experimental data of the type sometimes available to laboratory scientists. Our inability to use laboratory methods to independently vary treatments to eliminate or isolate spurious channels of causation places a fundamental limitation on the possibility of objective knowledge in the social sciences. In place of laboratory experimental variation, social scientists use subjective thought experiments. Assumptions replace data. In the jargon of modern econometrics, minimal identifying assumptions are invoked.

Because minimal identifying assumptions cannot be tested with data (all possible minimal assumptions for a model explain the observed data equally well, at least in large samples) and because empirical estimates of causal relationships are sensitive to these assumptions, inevitably there is scope for disagreement in the causal interpretation of social science data. Context, beliefs, and a priori theory resolve differences in causal interpretations. By definition, the available data cannot do so, although in principle, experiments can. The solution to the problem of causal inference lies outside of mathematical statistics and depends on contexts which are not universal."

James J Heckman and Richard Robb (1986)

Notation and Definition of Effects

The role of different variables and their relation to data

Mutually exclusive states of the world (treatments): $0,\ldots,M$

Potential outcomes (effects):	$Y^0, Y^1, \ldots, Y^M$	*M 'counterfactuals'; one observed*
Treatment status (causes):	$S \in \{0,1,\ldots,M\}$	exclusive states; *one observed*
Observed outcome:	$Y = \sum_{m=0}^{M} \underline{1}(S=m) Y^m$	
Observed characteristics:	X, Z	'exogenous'
Unobserved characteristics:	U	restrictions will be imposed

Pseudo experimental context: $Y^0,\ldots,Y^M \coprod S \mid X = x, U = u$

(independence in many cases not strictly necessary)

Popular (causal) effects

ATET: $\theta_0^{m,l} = E(Y^m \mid S = m) - E(Y^l \mid S = m)$

ATE: $\gamma_0^{m,l} = E(Y^m) - E(Y^l)$

- Effects need not be homogenous in the population
- Effects may vary over calendar time and over process time (distance to programme)
- Effects can be defined for **other reference populations** given by treatment status
- Effects can be **aggregated** with respect to treatments
- Effects can be **disaggregated** with respect to observable individual characteristics (*X*)
- Are these parameters policy relevant? Sometimes
- Major limitations of the model: $f(Y^m - Y^l)$, SUTVA and GE effects
- Roy (1951, *Oxford Economic Papers*), Rubin (1973, *Journal of Educational Psycho.*), Holland (1986, *JASA*)

Identification of causal effects

or *How to avoid selection bias* **?**

- Major source of selection bias is:

$$E(Y^l \mid S = m) \neq E(Y^l \mid S = l)$$

- Confounding factors that are correlated with **both** the outcomes and the selection

Two very different strategies

- If confounding factors are observable, then econometricians could control for them

- If confounding factors are **not** observable, other identification strategies are needed

- **External information needed for credible interpretation of any estimate**

Fundamental issues

- Learning from observable nonparticipation outcomes about unobservable counterfactual nonparticipation outcomes ? → Identification

- Identification depends on: *effects, assignment (selection) and outcome process, data*

- Different (untestable) minimal identifying restrictions could lead to different results!

- The optimal choice of estimator depends on:

  - implied minimum identifying restrictions

  - simplifying and variance reducing restrictions

  - data

- Surveys: Handbook of Labor Economics, III A: Angrist, Krueger (1999)
  Heckman, LaLonde, Smith (1999)

  Heckman, Robb (1985)

Almost perfect data (experiments)

- $E(Y^l \mid S = m; in\ experiment) = E(Y^l \mid S = l; in\ experiment)$ by randomization:

$$Y^0, ..., Y^M \coprod S \mid in\ experiment$$

➔ Jeff

- Randomization solves **most**, but **not all** identification problems
- Expensive !

Very informative data

Observe **all** characteristics (*X*) on which potential outcomes and selection depend **jointly**

Main identifying restriction (CIA: conditional independence assumption)

- $E(Y^l \mid S = m, X = x) = E(Y^l \mid S = l, X = x)$ by $Y^0,...,Y^M \coprod S \mid X = x, \forall x \in \chi$
- $E(Y^l \mid S = m) = E_X[E(Y^l \mid S = l, X = x) \mid S = m]$
- Selection bias can be avoided by controlling for specific observables

Further conditions:

- *X* must not be caused by the treatment!
- Common support: $0 < P(S = m \mid X = x) < 1, \quad \forall m \in \{0,...,M\}, \quad \forall x \in \chi$

Identified effects

ATE, ATET, ..., in common support

A note on estimation

- Estimation of $E_X[E(Y^l \mid S=l, X=x) \mid S=m]$ by parametric or nonparametric methods

- Match the distributions of X in the different states

- Practical problem: Dimension of conditioning set for $E_X[E(Y^l \mid S=l, X=x) \mid S=m]$

  Propensity score property:

$$Y^0,\ldots,Y^M \coprod S \mid X=x, \forall x \in \chi \quad \Rightarrow$$

$$Y^l \coprod S \mid [X=x, S \in \{m,l\}] \quad \forall x \in \chi \quad \Rightarrow$$

$$Y^l \coprod S \mid [P(S=m \mid S \in \{m,l\}, X=x), \quad S \in \{m,l\}] \quad \forall x \in \chi \quad \Rightarrow$$

$$\boxed{E(Y^l \mid S=m) = E_X\{E[Y^l \mid S=l, P^{m|m,l}(X=x)] \mid S=m\}} \quad \Rightarrow$$

If a *good* estimator of $P^{m|m,l}(X = x)$ available ➔ matching problem reduces to one dimension

Example

- Evaluation of Active Labour Market Policies in Switzerland (Gerfin, Lechner, 1999)
  - administrative records from unemployment insurance and pension system
  - data informative about employment history of last ten years
  - very detailed information about the near past
  - we should know everything what the labour office knows (and even more)
  - large and representative sample (20.000)
- Papers by Boris Augurzky, Thomas Brodaty, and Markus Frölich at this workshop

Practical considerations

- In multiple programmes with CIA: Pair-wise comparison are sufficient
- What *X* to select?
- *X* could include pre-treatment values of *Y* (be careful to check non causality condition!)
- If matching is on the propensity score: Choice of the correct probability model
- Distribution theory of matching estimators based on *estimated propensity score*
- Is it worth to condition on the propensity score ?
- Choice of estimator for conditional expectations of potential outcomes (if required)
- Major concern: Selection into programme may partially depend on expected gains (correlated with realised gains), that are at least partially unknown and unpredictable by the econometrician
- $E(Y^l \mid S = m, X = x) \neq E(Y^l \mid S = l, X = x)$, but bias may average out nevertheless

Papers

CIA: Rubin (1977, *Journal of Educational Statistics*)

Propensity score: Rosenbaum, Rubin (1983, *Biometrica*), Hahn (1998, *Econometrica*)

(New) Matching: Heckman, Ichimura, Todd (1998, *Review of Economic Studies*)

Multiple treatments: Imbens (1999, *NBER*), Lechner (1999, *Uni SG*)

Informative data, but something important is missing

$$Y^0,\ldots,Y^M \not\perp\!\!\!\perp S \mid X = x, \;\; \forall x \in \chi \quad \Rightarrow \quad E(Y^l \mid S = m, X = x) \neq E(Y^l \mid S = l, X = x)$$

$$Y^0,\ldots,Y^M \perp\!\!\!\perp S \mid X = x, U = u \quad \Rightarrow \quad E(Y^l \mid S = l, X = x, U = u) = E(Y^l \mid S = m, X = x, U = u)$$

General remarks

- CIA: $E_{X,U}[E(Y^l \mid S = l, X = x, U = u) \mid S = m] = E_X[E(Y^l \mid S = l, X = x) \mid S = m]$
- In the following: Find some other information (restriction) to be exploited

The mean of the counterfactual is the same before and after the treatment

Time is now used explicitly

$$Y_t^0, Y_t^1 \qquad t = \ldots, -1, 0, 1$$

$$\ldots, S_{-1} = 0, S_0 = 0 \; or \; 1, S_1 = 0, \ldots$$

Main identifying restriction (bias stability assumption)

- Means of potential outcomes do not change over time

$$E(Y_\tau \mid S_0 = 1) = E(Y_\tau^0 \mid S_0 = 1) = E(Y_t^0 \mid S_0 = 1) \;, \qquad \tau\,,\; \tau < 0\,,\; t\,,\; t > 0$$

- This is neither more general, nor more restrictive than CIA

Identified effect

- ATET

A note on estimation

- Before-after estimator

- *Longitudinal data:*

  (Mean for participants after treatment) - (mean for participants before treatment)

- *Repeated cross sections*:

  (Mean for participants after treatment) - (mean for future participants before treatment)

Practical considerations

- No need to use information of nonparticipants

- No need to observe pre- and post-treatment outcomes in same sample

- Not robust to changes over time (e.g. over the business cycle and over the life-cycle)

- Not robust to Ashenfelter's dip

The bias remains constant before and after treatment conditional on *X*

Main identifying restriction (bias stability assumption)

- The treatment has no impact in period τ, $\tau < 0$
- The bias is the same in at least one period τ, $\tau < 0$ and one period t, $t > 0$, $|\boldsymbol{X}$

$$E(Y_t^0 \mid S_0 = 1, X) - E(Y_t^0 \mid S_0 = 0, X) = E(Y_\tau^0 \mid S_0 = 1, X) - E(Y_\tau^0 \mid S_0 = 0, X)$$

$$[\text{or} \quad E(Y_t^0 - Y_\tau^0 \mid S_0 = 1, X) = E(Y_t^0 - Y_\tau^0 \mid S_0 = 0, X)] \quad \Rightarrow$$

$$E(Y_t^0 \mid S_0 = 1, X) = E(Y_t^0 \mid S_0 = 0, X) + [E(Y_\tau^0 \mid S_0 = 1, X) - E(Y_\tau^0 \mid S_0 = 0, X)] \quad \Rightarrow$$

$$E(Y_t^0 \mid S_0 = 1, X) = E(Y_t \mid S_0 = 0, X) + [E(Y_\tau \mid S_0 = 1, X) - E(Y_\tau \mid S_0 = 0, X)]$$

- Relation to CIA: $\boxed{(Y_t^0 - Y_\tau^0) \coprod S \mid X}$ vs. $\boxed{Y_t^0 \coprod S \mid X}$

  ➔ This is neither more general, nor is it weaker than CIA

Identified effect

- ATET (ATE)

A note on estimation

- Combines differencing over time with balancing distributions of X

- Combined with several other assumptions this model often justifies the use of panel estimators based on differencing (e.g. fixed effects); differences-in-differences

Example

- See paper by Martin Eichler later on

Literature

- Heckman and Hotz (1989, *JASA*), Heckman, Ichimura, Todd (1997, *REcSt*),

  Eichler, Lechner (2000, *Labour Economics*), Differences-in-differences-in-....-literature

Practical considerations

- Panel data not necessarily required

- Asymmetric changes in the labour market and individual behaviour over time

- What X to use ? Minimisation of bias $E(Y^0 \mid S=1, X=x) \neq E(Y^0 \mid S=0, X=x)$?

There is an instrument in the data

This is a very short version → *lots of other assumptions exist to define similar effects*

Main identifying ideas

- Exclusion restriction:

$$E[Y^0 \mid S(Z) = s(z), Z = z] = E[Y^0 \mid S(Z) = s(z)] \qquad \text{or}$$

$$E[Y^0 \mid S(Z) = s(z), X = x, Z = z] = E[Y^0 \mid S(Z) = s(z), X = x] \qquad s = 0,1 .$$

- Instruments influence selection:

$$P(S = s \mid Z = z) \neq P(S = s) \qquad \text{or} \qquad P(S = s \mid X = x, Z = z) \neq P(S = s \mid X = x)$$

- Note: Experiments are an extreme version of that

Identified effect

- **L**ocal**ATE** (for those who react when changing the value of the instrument *z*)

$$E[Y^1 - Y^0 \mid S(z') - S(z'') = 1] = \frac{E(Y \mid Z = z') - E(Y \mid Z = z'')}{P(S = 1 \mid Z = z') - P(S = 1 \mid Z = z'')}$$

- Is LATE something you want to know? The population depends on the instrument used
  - If the instrument defines a policy shift: Clear interpretation of LATE.
  - If not, it is not so clear what this should mean
  - More than 1 instrument, instrument has more than 2 values, continuous instrument: LIV
- LATE clarifies what is identified in standard IV without additional assumptions
- There is a considerable amount of disagreement in the field about IV and LATE

  ➔ The dust has not yet been settled in that area

A note on estimation

- Most simple if instrument is binary, and there are no covariates

- More difficult if adjustment for X is also necessary → matching methods could matter, or standard IV estimators (like 2SLS) if additional assumptions are imposed

Example

- Vietnam draft lottery

Literature

- Imbens, Angrist (1994, *Econometrica*), Angrist, Imbens, Rubin (1996, *JASA*), Heckman, Vytlacil (1999, *Prc. Natl. Sc.*)

Practical considerations

- Major problem: Is exclusion restriction plausible?

- Is LATE a relevant parameter given the specific instrument ?

Regression discontinuity design

Main identifying restriction

- Selection rule is known and deterministic: $P(S=1 \mid X > \tilde{x}) = 1,\ P(S=1 \mid X < \tilde{x}) = 0$

- Treatment effects are locally continuous in *x* at the cut-off $\tilde{x}$

- $Y^0, Y^1 \coprod S \mid X = x, \qquad x \in \{\tilde{x} \pm \varepsilon\} \qquad \varepsilon \to 0 \qquad$ (local CIA)

$$\Rightarrow \pi(\tilde{x}, \varepsilon) = E(\underbrace{Y^1}_{Y} \mid X = \tilde{x} + |\varepsilon|) - E(\underbrace{Y^0}_{Y} \mid X = \tilde{x} - |\varepsilon|)$$

$$\Rightarrow \lim_{\varepsilon \to 0} \pi(\tilde{x}, \varepsilon) = E(Y^1 \mid X = \tilde{x}) - E(Y^0 \mid X = \tilde{x}) = \gamma_0^{1,0}(\tilde{x})$$

- In a local neighbourhood to the threshold, selection is like a random event

Identified effect

- Only effects local to the cut-off are nonparametrically identified

A note on estimation

- Nonparametric estimation close to cut-off points
- Regression type estimator if additional assumptions are used

Example

- Class size by Angrist and Lavy (1996, Hebrew Univ.)
- Anti-Discrimination laws by Hahn, Todd, van der Klaaw (1999)

Literature

- Thristlewaite, Campbell (1960, *Journal of Educational Psychology*)
- Angrist, Lavy (1996), Hahn, Todd, van der Klaaw (1999, *NBER*)

Practical considerations

- This approach has not (yet) been much in use
- Appears to be promising in many cases when there is no instrument and no CIA
- No common support at all!
- What is local ?
- Fuzzy selection rules introduce considerable complications

We know (part of) the joint distribution of *U* and its connection to the potential outcome and selection variables: The control function approach to selectivity correction

Main identifying restriction

- A control function takes care of the selection bias
- Extreme variety of specific assumptions. In many cases the following is assumed:
  - Index sufficiency
  - Exclusion restriction
  - Functional forms of ...

Identified effect

- Everything can be identified
- However, in general models there may be problems

A note on estimation

- Easy for very restrictive models (like Heckman, 1979)

- Generalisations (like semiparametric selection models) come at a cost

Example

- Millions (?) of papers using the Heckman 2-step method (Heckit)

Literature

- Heckman (1979, *Econometrica*)

- Literature on semiparametric selection models, e.g. Andrews, Schafgans (1998, *REcSt*), Heckman (1990, *AER*), Survey by Vella (1998, *JHR*)

Practical considerations

- Credibility problem

- Simple models are overly restrictive, complex models are still hard to estimate

We know very little → Bounds

Main identifying restriction

- $Y^0,...,Y^M$ have limited support

  Unknown quantities: $E(Y^m \mid S=l)$ and $E(Y^l \mid S=m)$

  Largest / smallest possible value of Y^m in group *l* and Y^l in group *m* known

  ➔ Upper and lower bounds on $E(Y^m \mid S=l)$ and $E(Y^l \mid S=m)$

  ➔ Upper and lower bounds for ATE and ATET without any restriction on selection process

- Conditioning on *X* without any (unusual) problem
- Additional a priori information / restrictions could be included (not always easy)

Identified effect

- Bounds, no point estimate

A note on estimation

- Estimate $E(Y^m \mid S = m)$ and $E(Y^l \mid S = l)$ nonparametrically if sample size permits
- Taking account of sampling error in the estimation of the bounds may be very difficult

Example

- Lechner (1999, *Econometrics Journal*): Effects of training in East Germany using a large but not very informative sample (German microcensus)

Literature

- Econometrics: Manski (1990, *AER*)
- Applications: Manski, Sandefur, McLanahan, Powers (1992, *JASA*), Lechner ('99, *EctrJ*)
- → IV: Hotz, Mullin, Sanders (1997, *REcSt*)

Practical considerations

- Are the bounds informative? In many cases they are not

Really bad data

better forget it

and convince somebody to

collect better data !

Conclusions

- Observational studies can give valuable insight, if identification is credible
- Observational studies can be much cheaper than experiments
- Typically identification is less credible than in experiments
- There appears to be a trend away from identification by complex statistical modelling
  Advantage of this change: Improved communication with non-econometricians
- Better data will lead to more credible identification and hence to better results
- Better estimators allow to use the potential of the given data fully
- **There is nothing like a method that always works! The method of choice depends on the data, on the complexity of the problem, and on prior information**
- Cross-sectional estimators and panel estimators (or models)
- Natural experiments

Important open questions for practical applications

- Endogenous timing of the start of the programme
- Timing and the choice of different subprograms in large policies
- Programme careers
- Should the time in the programme count towards the effect?
- What are the most interesting outcome variables?

- The connection between the micro and the macro level of evaluations (GE effects)

Put everything into the perspective of 'The History of Econometrics'

Heckman, James J. (1999): „Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective“, *NBER Working Paper*, 7333