Interpreting IV, Part 1

James J. Heckman
University of Chicago
Extract from: Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy
James Heckman (JEL 2010)

Econ 312, Spring 2021

- Roy (1951): Agents face two potential outcomes (Y_0, Y_1) with distribution $F_{Y_0,Y_1}(y_0, y_1)$ where "0" refers to a no treatment state and "1" refers to the treated state and (y_0, y_1) are particular values of random variables (Y_0, Y_1).
- More generally, set of potential outcomes is $\{Y_s\}_{s \in \mathcal{S}}$ where $\mathcal{S}$ is the set of indices of potential outcomes.
- Roy model $\mathcal{S} = \{0, 1\}$.

- Analysts observe either Y_0 or Y_1, but not both, for any person.
- In the program evaluation literature, this is called the **evaluation problem**.

- The **selection problem**.
- Values of Y_0 or Y_1 that are observed are not necessarily a random sample of the potential Y_0 or Y_1 distributions.
- In the original Roy model, an agent selects into sector 1 if $Y_1 > Y_0$.

$$D = \mathbf{1}(Y_1 > Y_0), \tag{1}$$

- Generalized Roy model (C is the cost of going from "0" to "1")

$$D = \mathbf{1}(Y_1 - Y_0 - C > 0). \tag{2}$$

- The outcome observed for any person, Y, can be written as

$$Y = DY_1 + (1 - D)Y_0. \tag{3}$$

- $\mathcal{I}$ denotes agent information set.
- In advance of participation, the agent may be uncertain about all components of (Y_0, Y_1, C).
- Expected benefit: $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I})$.
- Then

$$D = \mathbf{1}(I_D > 0). \tag{4}$$

- The decision maker selecting "treatment" may be different than the person who experiences the outcomes (Y_0, Y_1).

- The *ex-post* objective outcomes are (Y_0, Y_1).
- The *ex-ante* outcomes are $E(Y_0 \mid \mathcal{I})$ and $E(Y_1 \mid \mathcal{I})$.
- The *ex-ante* subjective evaluation is I_D.
- The *ex-post* subjective evaluation is $Y_1 - Y_0 - C$.
- Agents may regret their choices because realizations may differ from anticipations.

- $Y_1 - Y_0$ is the individual level treatment effect.
- Also, the Marshallian ceteris paribus causal effect.
- Because of the evaluation problem, it is generally impossible to identify individual level treatment effects.
- Even if it were possible, $Y_1 - Y_0$ does not reveal the *ex-ante* subjective evaluation I_D or the *ex-post* assessment $Y_1 - Y_0 - C$.

- Economic policies can operate through changing (Y_0, Y_1) or through changing C.

Population Parameters of Interest

- Conventional parameters include the Average Treatment Effect (ATE $= E(Y_1 - Y_0)$), the effect of Treatment on The Treated (TT $= E(Y_1 - Y_0 \mid D = 1)$), or the effect of Treatment on the Untreated (TUT $= E(Y_1 - Y_0 \mid D = 0)$).

THE UNIVERSITY OF CHICAGO

- In positive political economy, the fraction of the population that perceives a benefit from treatment is of interest and is called the **voting criterion** and is

$$\Pr(I_D > 0) = \Pr(E(Y_1 - Y_0 - C \mid \mathcal{I}) > 0).$$

- In measuring support for a policy in place, the percentage of the population that *ex-post* perceives a benefit is also of interest: $\Pr(Y_1 - Y_0 - C > 0)$.

- Determining marginal returns to a policy is a central goal of economic analysis.
- In the generalized Roy model, the margin is specified by people who are indifferent between "1" and "0", i.e., those for whom $I_D = 0$.
- The mean effect of treatment for those at the margin of indifference is

$$E(Y_1 - Y_0 \mid I_D = 0).$$

Treatment Effects Versus Policy Effects

- Policy Relevant Treatment Effect (Heckman and Vytlacil, 2001) extends the Average Treatment Effect by accounting for voluntary participation in programs.
- "b": baseline policy ("before") and "a" represent a policy being evaluated ("after").
- Y^a: outcome under policy a; Y^b is the outcome under the baseline.
- (Y_0^a, Y_1^a, C^a) and (Y_0^b, Y_1^b, C^b) are outcomes under the two policy regimes.

- If some parameters are invariant to policy changes, they can be safely transported to different policy environments.
- Structural econometricians search for policy invariant "deep parameters" that can be used to forecast policy changes.

- Under one commonly invoked form of policy invariance, policies keep the potential outcomes unchanged for each person: $Y_0^a = Y_0^b$, $Y_1^a = Y_1^b$, but affect costs ($C^a \neq C^b$).
- Such invariance rules out social effects including peer effects and general equilibrium effects.

- Let D^a and D^b be the choice taken under each policy regime.
- Invoking invariance of potential outcomes, the observed outcomes under each policy regime are $Y^a = Y_0 D^a + Y_1(1 - D^a)$ and $Y^b = Y_0 D^b + (1 - D^b)$.

- The **Policy Relevant Treatment Effect** (PRTE) is

$$\text{PRTE} = E(Y^a - Y^b).$$

- Comparison of aggregate outcomes under policies "*a*" and "*b*". PRTE extends ATE by recognizing that policies affect incentives to participate (C) but do not force people to participate.
- Only if C is very large under b and very small under a, so there is universal nonparticipation under b and universal participation under a, would ATE and PRTE be the same parameter.

Link to Appendix

Appendix

Proof

(Keep X implicit)

$$
\begin{aligned}
E\left(Y_p\right) &= \int_0^1 E(Y_p \mid P_p(Z_p) = t)\, dF_{P_p}(t) \\
&= \int_0^1 \Big[\int_0^1 [\mathbf{1}_{[0,t]}(u_D) E(Y_{1,p} \mid U_D = u_D) \\
&\qquad + \mathbf{1}_{(t,1]}(u_D) E(Y_{0,p} \mid U_D = u_D)]\, du_D \Big]\ dF_{P_p}(t) \\
&= \int_0^1 \Big[\int_0^1 [\mathbf{1}_{[u_D,1]}(t) E(Y_{1,p} \mid U_D = u_D) \\
&\qquad + \mathbf{1}_{(0,u_D]}(t) E(Y_{0,p} \mid U_D = u_D)]\, dF_{P_p}(t) \Big]\ du_D \\
&= \int_0^1 \big[(1 - F_{P_p}(u_D)) E(Y_{1,p} \mid U_D = u_D) \\
&\qquad + F_{P_p|X}(u_D) E(Y_{0,p} \mid U_D = u_D) \big]\ du_D .
\end{aligned}
$$

Proof

- Comparing policy p to policy p',

$$E\left(Y_p\right) - E\left(Y_{p'}\right) = \int_0^1 \underbrace{E(Y_1 - Y_0 \mid U_D = u_D)}_{\text{MTE}(u_D)} (F_{P_{p'}}(u_D) - F_{P_p}(u_D))\, du_D,$$

which gives the required weights.

- Policies shift the distribution of $P(Z)$.
- They keep the distribution of Y_1 and Y_0 unchanged.

Proof

- This derivation involves changing the order of integration.
- Note that from finiteness of the mean,

$$E\left|\mathbf{1}_{[0,t]}(u_D)E(Y_{1,p} \mid U_D = u_D) + \mathbf{1}_{(t,1]}(u_D)E(Y_{0,p} \mid U_D = u_D)\right|$$
$$\leq E(|Y_1| + |Y_0|) < \infty,$$

$\therefore$ the change in the order of integration is valid by Fubini's theorem.