Econometric Causality

(2008) *International Statistical Review*, 76(1):1-27

James J. Heckman

Spencer/INET Conference
University of Chicago

Econometric Causality

- The econometric approach to causality develops explicit models of outcomes where the causes of effects are investigated and the mechanisms governing the choice of treatment are analyzed.
- The relationship between treatment outcomes and treatment choice mechanisms is studied.
- A careful accounting of the unobservables in outcome and treatment choice equations facilitates the design of estimators to solve selection and evaluation problems.
- It also facilitates understanding of the **causal mechanisms** by which outcomes are produced: both outcome equations and treatment assignment (choice) equations.

- Both objective and subjective evaluations are considered, where subjective valuations are those of the person receiving treatment as well as the persons assigning it.
- Differences between anticipated and realized objective and subjective outcomes are analyzed.
- Models for simultaneous treatment effects are developed. (Joint causation)
- A careful distinction is made between models for potential outcomes and empirical methods for identifying treatment effects.

- **The unaided "treatment effect" (Neyman-Rubin) model is not the appropriate framework for addressing the causal influence of personality on outcomes.**
- The treatment effect model focuses on "effects of causes" not causes of effects.
- **The econometric approach examines the "causes of the effects" and the mechanisms that produce outcomes in order to consider and evaluate effective interventions that promote personality.**

- An example of a structural relationship (Haavelmo, 1944, *Econometrica*)

$$Y = X_b\beta_b + X_p\beta_p + U \qquad (*)$$

U: A variable unobserved *by the analyst*
X_b: background variables
X_p: policy variables (can manipulate by interventions)
$*$ is an "all causes" model
External manipulations define causal parameters:
Variations in (X_b, X_p) that hold U fixed
If the coefficients (β_b, β_p) are invariant to shifts in (X_b, X_p), then $(*)$ is structural.

- Notice that OLS is

$$E^*(Y \mid X_b, X_p) = X_b\beta_b + X_p\beta_p + E^*(U \mid X_b, X_p)$$

  where E^* is a linear projection.
- OLS is not estimating a structural relationship.
- If $E(U \mid X_b, X_p) = 0$, OLS gives a structural estimator for (β_b, β_p).

- If

$$E^*(U \mid X_b, X_p) = E^*(U \mid X_b)$$

  and the coefficient in the original model is invariant to manipulations in X_p then OLS is structural for β_p.

- But not necessarily for β_b.

The Structural Versus the Program Evaluation Approach to Evaluating Economic Policies

- Causality at the individual level.
- Based on the notion of controlled variation — variation in treatment holding other factors constant.
- This is Alfred Marshall's (1890) *ceteris paribus* clause which has been the operational definition of causality in economics for over a century.
- It is distinct from other notions of causality sometimes used in economics that are based on prediction (e.g., Granger, 1969, and Sims, 1972).

- Two distinct tasks in causal inference and policy analysis: (a) Defining counterfactuals and (b) Identifying causal models from data.
- Table 1 delineates the two distinct problems.

Table 1: Two Distinct Tasks that Arise in the Analysis of Causal Models

Task	Description	Requirements
1	Defining the Set of Hypotheticals or Counterfactuals	A Well-specified Scientific Theory
2	Identifying Causal Parameters from Data	Mathematical Analysis of Point or Set Identification Joined With Estimation and Testing Theory

Policy Evaluation Problems and Criteria of Interest

P1

Evaluating the Impacts of Implemented Interventions on Outcomes Including Their Impacts on the Well-Being of the Treated and Society at Large.

- Objective evaluations
- Subjective evaluations
- Ex ante and ex post

P2

Forecasting the Impacts (Constructing Counterfactual States) of Interventions Implemented in One Environment in Other Environments, Including Impacts on Well-Being.

- This is the problem of *external validity*: taking a treatment parameter or a set of parameters identified in one environment to another environment.

P3

Forecasting the Impacts of Interventions (Constructing Counterfactual States Associated with Interventions) Never Historically Experienced, Including Their Impacts on Well-Being.

- **P3** is a problem that policy analysts have to solve daily.
- Structural econometrics addresses this question.
- The program evaluation approach does not except through "demonstration programs" (i.e. that explicitly implement the policies).

A Prototypical Economic Model for Causal Analysis, Policy Evaluation and Forecasting the Effects of New Policies

- 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 $\mathcal{S} = \{0, 1\}$.
- The Y_0, Y_1 depend on $X = (X_b, X_p)$,

  e.g., $$E(Y_0 \mid X) = \mu_0(X)$$ $$E(Y_1 \mid X) = \mu_1(X)$$

- 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

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

- C can depend on cost shifters (e.g., Z)

$$E(C \mid Z) = \mu_C(Z)$$

- $\mathcal{I}$ denotes agent information set **of the agent**.
- 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 (Task 2).
- 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.
- Changes in $Y_0, Y_1,$ and C can be brought about by changing both the X and the Z.
- The structural approach considers policies affecting both returns and costs.

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

- 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.
- Designed to address problems **P2** and **P3**.
- "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.

- Policy invariance facilitates the job of answering problems **P2** and **P3**.
- 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

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

- Benthamite 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.

The Econometric Approach Versus the "Rubin" Model Treatment Effect Approach

- Econometric approach examines the causes of effects
- How Y_1 and Y_0 vary as X varies
- How treatment (D) gets determined through variations in Z
- This is the goal of science
- The treatment effect approach ("Rubin model") looks at *effects of causes*
- Does not investigate **mechanisms** of causation
- Framework is ill-suited to the study of personality psychology where causal mechanisms need to be developed

Table 2: Comparison of the Aspects of Evaluating Social Policies that are Covered by the Neyman-Rubin Approach and the Structural Approach

	Neyman-Rubin Framework	Structural Framework
Counterfactuals for objective outcomes (Y_0, Y_1)	Yes	Yes
Agent valuations of subjective outcomes (I_D)	No (choice-mechanism implicit)	Yes
Models for the causes of potential outcomes	No	Yes
Ex ante versus *ex post* counterfactuals	No	Yes
Treatment assignment rules that recognize voluntary nature of participation	No	Yes
Social interactions, general equilibrium effects and contagion	No (assumed away)	Yes (modeled)
Internal validity (problem **P1**)	Yes	Yes
External validity (problem **P2**)	No	Yes
Forecasting effects of new policies (problem **P3**)	No	Yes
Distributional treatment effects	No[a]	Yes (for the general case)
Analyze relationship between outcomes and choice equations	No (implicit)	Yes (explicit)

[a]An exception is the special case of common ranks of individuals across counterfactual states: "rank invariance." See the discussion in Abbring and Heckman (2007).

Methods of Estimation (Task 2)

- Rubin-Neyman model elevates randomization to the "gold standard" — it is not.
- After explicating the "Rubin model," Holland makes a very revealing claim: there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.
- This statement confuses the act of defining a causal effect (a purely mental act performed within a model) with empirical difficulties in estimating it.
- It confuses the tasks of formulating a theory and the concept of causality within a model with the practical problems of testing it and estimating the parameters of it.

- Unaided, data from randomized trials cannot identify the voting criterion ($\Pr(Y_1 - Y_0) > 0$) i.e. percentage of people who benefit.
- Matching assumes that the marginal recipient of treatment gets the same return as the average.
- Unaided IV or "LATE" identifies people at an unspecified margin — doesn't tell us which people are induced to switch.