Causality frameworks

Miguel Olivo-Villabrille

Causality

- Here we do a brief review of causality (also known as causal inference or causal analysis).
  - This will be a high-level overview.
- Establishing causality is the ‘killer app’ of empirical economists.
  - As opposed to other data disciplines that often times are content with simply obtaining correlations.

Causality

- We say that X causes Y if we were to intervene and change X without changing anything else and Y would change as a result.
- Many interesting questions we are interested in answering with data are causal.
- Of course, some questions are non-causal. For example, questions about prediction (machine learning excels in this).
- Almost every 'why' question is causal.
- We aim to answer the question 'how do we know if X causes Y?'

Causality

- Some causal relationships:
  - Setting a light switch to on causes the light to be on.
  - Pushing the break pedal causes a car to slow down.
  - Obtaining a college degree increases earnings.
  - Exercise improves overall health.
- Correlation is not causation:
  - People carrying umbrellas on rainy days (even before it rains) [causality goes in the other direction].
  - Consumption of wine and longevity.

Causality

- 'X causes Y' doesn't necessarily mean that X is the only thing that causes Y.
- It does not mean that all the change on Y is due to X.
- The important thing is that at least X changes the probability of Y occurring.

Causality

- But why do we need a causal framework?
- Statistics alone is not sufficient to establish causality.
- Statistics lacks 'directionality.'
- In statistics, two random variables X, Y can be fully described by their joint distribution.
- However, causality requires additional information regarding the direction of the causal relationship between X and Y.
  - For example, a (purely statistical) regression of Y on X with highly significant coefficients tells us nothing about the direction of causality, since regressing X on Y will also produce significant coefficients.

Causality

- We will briefly see three frameworks for establishing causality.
- These frameworks are commonly employed in economics and public policy.
- Potential Outcomes Framework.
- Structural Causal Model.
- ‘Econometric’ model of causality.

Potential Outcomes Framework

- Also called the Rubin causal model, Neyman-Rubin causal model.
- The Potential Outcomes Framework was introduced by Jerzy Neyman in 1923 for randomized experiments.
- Extended to observational data by Donald Rubin in 1974-1980.
- It is now the basic framework for much of modern statistical analysis aiming to establish causality.
- Applications abound: labor econ, health econ, family econ, environmental econ, industrial organization, corporate finance, trade, agricultural econ, sociology, political science, psychology, etc.
- 2021 Nobel in Econ awarded to Card, Angrist & Imbens: all empirical economists, citation for Angrist & Imbens mentions causality.

Potential Outcomes Framework

- Some references:
- Mastering 'Metrics: The Path from Cause to Effect, by Joshua Angrist and Jörn-Steffen Pischke, 2014.
- Causal Inference: The Mixtape, by Scott Cunningham, 2021.
- Mostly Harmless Econometrics: An Empiricist's Companion, by Joshua Angrist and Jörn-Steffen Pischke, 2009.
- Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, by Guido Imbens & Donald Rubin, 2015.
- Endogeneity in Empirical Corporate Finance, by Michael R. Roberts & Toni M. Whited.

Potential Outcomes Framework

- Four basic strategies: randomized trials, matching, instrumental variables (includes regression discontinuity), and difference-in-differences.
- We now present an overview of its basic components.
- Y is the outcome (for example wages), D is the treatment or intervention (for example university graduation).
  - Treatment does not have to be binary, but for simplicity let's assume it is.
- The model describing the impact of treatment on outcome is

$$Y = h(D, U).$$

- U is an individual-level unobserved random factor (for example individual abilities, skills, work ethic, interpersonal connections, preferences, etc).
  - Often, instead of writing U, authors prefer to index by i: $Y_i = h_i(D_i)$.

Potential Outcomes Framework

- We can simplify notation further, by defining $Y(0) = h(0, U)$, and $Y(1) = h(1, U)$.
- We do not observe both $Y(1)$ and $Y(0)$ for a single individual.
  - These are called potential outcomes of treatment and non-treatment.
- We only observe the realized value $Y = Y(0) + D(Y(1) - Y(0))$.
  - If $D = 1$ (individual is treated) then $Y = Y(1)$ (the observed outcome is the potential outcome of treatment).
  - If $D = 0$ (individual is not treated) then $Y = Y(0)$ (the observed outcome is the potential outcome of non-treatment).

Potential Outcomes Framework

- We can then define the causal or treatment effect of D on Y: The causal effect of D on Y is

$$TE(U) = Y(1) - Y(0) = h(1, U) - h(0, U)$$

  the change in Y due to treatment while holding U constant.
- We can also define the average treatment effect: The average causal effect of D on Y is

$$ATE = \mathbb{E}[TE(U)] = \int_{\mathcal{U}} TE(u) f(u)\, \mathrm{d}u$$

  where $f(u)$ is the density of U, and the integration is over the domain of f, $\mathcal{U}$.

Potential Outcomes Framework

- We don't observe both $Y(1)$ and $Y(0)$, can we compare observable outcomes for treated and untreated individuals?

$$\mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0] = \int_{\mathcal{U}} h(1, u) f(u \mid D = 1) \, \mathrm{d}u - \int_{\mathcal{U}} h(0, u) f(u \mid D = 0) \, \mathrm{d}u \, .$$

- In general, that comparison does not yield the ATE, unless D and U are independent.
- If D is randomly assigned (as in a randomized trial) then D and U become independent and we have $f(u \mid D) = f(u)$, so:

$$\mathbb{E}[Y \mid D = 1] - \mathbb{E}[Y \mid D = 0] = \int_{\mathcal{U}} (h(1, u) - h(0, u)) f(u) \, \mathrm{d}u = ATE.$$

Potential Outcomes Framework

- The framework can be extended to allow for covariates X:

$$Y = h(D, X, U).$$

- The causal or treatment effect of D on Y now becomes

$$TE(X, U) = h(1, X, U) - h(0, X, U)$$

  the change in Y due to treatment holding X and U constant.
- The conditional average causal effect of D on Y conditional on $X = x$ is

$$ATE(x) = \mathbb{E}[TE(X, U) \mid X = x] = \int_{\mathcal{U}} TE(x, u) f(u \mid x) \, \mathrm{d}u$$

  where $f(u \mid x)$ is the conditional density of U given $X = x$.
- The unconditional average causal effect of D on Y is

$$ATE = \mathbb{E}[ATE(X, U)] = \int_{\mathcal{X}} ATE(x) f(x) \, \mathrm{d}x$$

  where $f(x)$ is the density of X, and $\mathcal{X}$ is the domain of X.

Potential Outcomes Framework

- Rather than requiring that D and U be independent, we can instead require that D and U be independent conditional on X.
- Conditional Independence Assumption (CIA):
  - Conditional on X, the random variables D and U are independent.
- Then we have, $f(u \mid D, X) = f(u \mid X)$, so

$$\begin{aligned}
&\mathbb{E}[Y \mid D = 1, X = x] - \mathbb{E}[Y \mid D = 0, X = x] \\
&\quad = \int_{\mathcal{U}} h(1, x, u) f(u \mid x)\,\mathrm{d}u - \int_{\mathcal{U}} h(0, x, u) f(u \mid x)\,\mathrm{d}u \\
&\quad = \int_{\mathcal{U}} TE(x, u) f(u \mid x)\,\mathrm{d}u = ATE(x).
\end{aligned}$$

- In the context of linear regression, we have $Y = \beta_1 + \beta_2 D + \beta_3 X + u$, then $ATE = \beta_2$.

Potential Outcomes Framework

- That basic framework can be applied to all research designs in the Rubin causal framework.
  - IV extends the framework to accommodate the relationship between treatment and instrument.
- Notice we need some kind of independence between treatment D and U (randomized trials or natural experiments).
  - Except for difference in differences, which instead relies on parallel trends assumption. Still, randomization buys a lot of credibility in DiD.
- Some implicit assumptions of the framework:
  - Stable unit treatment value assumption (SUTVA): rules out spillover effects, network effects, general equilibrium effects.

Structural Causal Model

- Also called Do-calculus.
- The Structural Causal Model was introduced by Judea Pearl in 1995-2000.
- Introduces an operator to indicate causality $\Pr(Y \mid \mathrm{do}\,(X))$.
- Perhaps the most defining feature of the framework is its use of Directed Acyclic Graphs (DAG).
- Although earlier attempts at using graphs for causal analysis date from 1920s with the work of Sewall Wright.
- SCM is very popular in computer science and artificial intelligence communities.
- Economists have been slow in adopting this framework, but there are some examples of applications (Spiegler, 2020).
- Interestingly, Imbens is skeptical about SCM (Imbens, 2020).

Structural Causal Model

- Some references:
- Causality, by Judea Pearl, 2009.
- Causal Inference in Statistics: A Primer, by Judea Pearl, Madelyn Glymour & Nicholas Jewell, 2016.
- Counterfactuals and Causal Inference: Methods and Principles for Social Research, by Stephen Morgan & Christopher Winship, 2014.
- The Effect: An Introduction to Research Design and Causality, by Nick Huntington-Klein, 2022.
- The Book of Why, by Judea Pearl & Dana Mackenzie, 2018.
- Online tool: `dagitty.net`. It is also an R package.

Structural Causal Model

- A DAG is a collection of vertices (or nodes) and edges that satisfies certain conditions.
  - Edges are directed: they go from one node into another, the direction is indicated by arrows.
    - The direction of the arrow represents the direction of causality.
  - There are no cycles: there isn't a directed path from any node into itself.

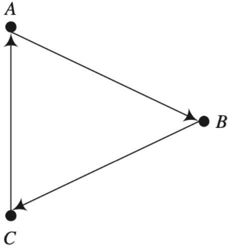


Figure 3.1 A directed graph that includes a cycle.

Structural Causal Model

- Some useful terminology and conventions:
  - Each node represents an observed or unobserved random variable.
  - Observed variables are represented by solid circles •, unobserved variables by hollow circles ◦.
  - All observed variables are assumed to be influenced by unobserved random variables. This is often omitted from causal graphs for the sake of simplicity.
    - If two variables are caused by a common unobserved variable, it is often typical to join those variables by a bidirectional dashed edge.

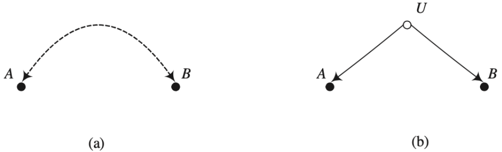
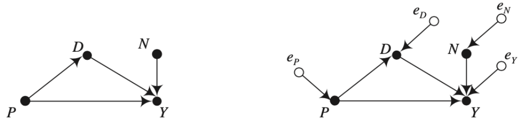


Figure 3.2 Two representations of the joint dependence of A and B on unobserved common causes.

Structural Causal Model

- More terminology:
  - A path is any sequence of edges pointing in any direction that connects one variable to another.
  - A directed path is a path in which all edges point in the same direction.
  - A variable is a descendant of another variable if it can be reached by a directed path.
  - For directed paths of length one, as in $A \rightarrow B$, the variable A is the parent while the variable B is the child.
  - A descendant of a variable is another variable than can be reached by a directed path from the first variable.
  - In a DAG, no variable is a descendant of itself.



(a) Standard representation (b) Under magnification

Figure 3.7 Equivalent directed graph representations of the effects of parental background (P), charter schools (D), and neighborhoods (N) on test scores (Y).

Structural Causal Model

- Why go through all that trouble? Econometrics is complicated enough already.
- It turns out that there is a one-to-one mapping between graphs and systems of structural equations.
- Moreover, by following a given set of rules, identification of causal effects can be assessed from causal graphs.
- Let's see an example: Figure 3.7 in the previous slide corresponds to the structural equations:

$$\begin{aligned} P =& f_P(e_P) \\ D =& f_D(P, e_D) \\ N =& f_N(e_N) \\ Y =& f_Y(D, P, N, e_Y). \end{aligned}$$

- These are nonparametrically specified, but we can think of them as being linear regressions, ie $Y = \beta_0 + \beta_1 D + \beta_2 P + \beta_3 N + e_Y$.

Structural Causal Model

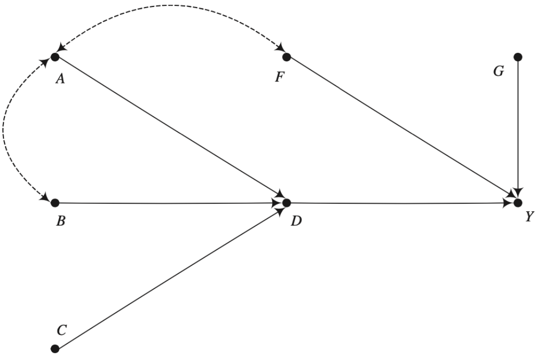
- This is not only useful for systems of equations, but for single equation models as well.
- We can assess whether a causal effect can be identified and under what conditions (for example whether it is necessary to control or condition for other variables) by examining a graph.
- Following the previous example, those equations correspond to the regression:

  $$Y = \beta_0 + \beta_1 D + \beta_2 P + \beta_3 N + e_Y,$$

  where D is correlated with P.
- The question is whether this regression identifies the causal effect of D on Y.
- Turns out that (using the rules provided by the SCM) as long as we control for P, the causal effect of D on Y is identified.
- Moreover, controlling for N is not necessary for identification of the causal effect!
  - This is a because D and N are independent, but D and P are not independent.

Structural Causal Model

- A final example:



- In this graph, controlling for F is sufficient to identify the causal effect of D on Y (there are also other strategies available).
- Much easier and faster to point out what variables we need to control for, and which variables are unnecessary to control for.

Structural Causal Model

- The Structural Causal Model is much richer than what I showed here.
- In fact we didn't discuss one of its more fundamental innovations, the do-operator.
- It is also worth pointing out that this is a complete system, in the sense that it is possible to discover new identification strategies from iteratively applying its rules to a novel estimation problem.
- Things that the SCM has clarified include:
  - Bad controls.
  - Berkson's paradox.

Econometric Model of Causality

- 'Econometric' model of causality, pioneered by Ragnar Frisch & Trygve Haavelmo, and formalized by James Heckman & Rodrigo Pinto.
- It is the causality framework used in structural econometrics.
  - Structural econometrics aims to estimate deep parameters of economic models (policy-invariant parameters).
- Structural econometrics is prevalent in fields like IO, that model strategic interactions between firms in an industry.
- Also used in fields in which it is important to model and estimate policy-invariant parameters, for example in health and education, where it is important to estimate parameters of the demand function.
- Also in macroeconomics, where researchers employ general equilibrium models to forecast the economy.
- Nowadays it is also used in the private sector, for example in auctions (eg eBay), and for real-time pricing (eg Uber).

Econometric Model of Causality

- Some references:
- Econometric Causality (2008), by Heckman.
- Causal Analysis After Haavelmo (2015), by Heckman & Pinto.
- Causality and Econometrics (2022), by Heckman & Pinto.
- The Econometric Model for Causal Policy Analysis (2022), by Heckman & Pinto.
- Econometric causality: The central role of thought experiments (2024), by Heckman & Pinto.
- Some references for structural econometrics:
  - Reiss & Wolak, 2007.
  - Galiani & Pantano, 2022.
  - Low & Meghir, 2017.
  - Keane, 2010a.
  - Keane, 2010b.

Econometric Model of Causality

- The Econometric Model of Causality is more general than both of the frameworks we saw before.
  - All treatment effect parameters available in RCM and SCM can be recasted in the EMC.
  - Allows for more comprehensive analysis of phenomena.
- It relies on a hypothetical model to perform counterfactual experiments that are used to assess causality.
- More importantly, it can be used to establish causality in situations that are pervasive in economics and where the other two frameworks fail to consider:
  - Social and strategic interactions.
  - Peer effects.
  - General equilibrium effects.
  - It also allows for identification strategies that make use of functional form restrictions.

Econometric Model of Causality

- Illustrating with a simple example.
- Consider a family that must decide how much of wife's labour to provide and how much hours of formal childcare to use.
- The family makes those decisions by maximising family utility subject to a budget constraint.

$$\max_{h,c} U(y, \ell, X; \beta) \quad \text{subject to}$$

$$y \leq \tau(y_0 + wh, X) - \psi(pc, y_0 + wh, X).$$

- Decision variables: c is hours of childcare, h is hours of work of wife,
- y is consumption, y_0 is exogenous income, ℓ is leisure time of wife, w is wage rate of wife, p is price of childcare, X is vector of demographic variables.
- Function τ encapsulates tax system, function ψ encapsulates childcare subsidy.
- β is a vector of parameters to be estimated.

Econometric Model of Causality

- For simplicity, we consider h and c to be binary variables.
- $h \in H = \{0, 35\}$, corresponding to no work, full-time work.
- $c \in C = \{0, 35\}$, corresponding to no childcare, full-time childcare.
- A more realistic model would consider restrictions between the choices of h and c, but we refrain from that here for simplicity.
- The choices of h and c can be combined into 4 pairs $(h, c) \in H \times C = \{(0, 0), (0, 35), (35, 0), (35, 35)\}$.
- The framework then boils down to a multinomial choice model.

Econometric Model of Causality

- Suppose we have estimated the model and have estimates of β, denoted $\widehat{\beta}$.
- Now consider a new policy that results in a change in the price of childcare p, the new price is denoted $\widetilde{p}$.
- If $\widehat{\beta}$ are policy-invariant (they are 'deep' parameters), we can then solve for the family choices of h and c resulting from that new policy.
- The hypothetical model is the same model but we replace p by $\widetilde{p}$, and use the estimates $\widehat{\beta}$.

$$\max_{h,c} U(y, \ell, X; \widehat{\beta}) \quad \text{subject to}$$

$$y \leq \tau(y_0 + wh, X) - \psi(\widetilde{p}c, y_0 + wh, X).$$

- We then have counterfactual or predicted choices under the new policy.
- Within the model, $\widetilde{p}$ causes families to change their choices. We can then analyse those choices in a variety of ways (costings, welfare, usage, etc).

The purpose of econometric policy evaluation

- Why do we do all this?
- Good policy analysis is causal analysis.
- Good causal analysis yields policy levers that can be pulled to achieve a policy goal.
- Good causal analysis evaluates the impacts of policy interventions at small and large scale.
  - This is particularly relevant for public servants, who frequently are asked to answer questions about costs and impacts of policies.

The four classes of policy problems

- Heckman proposes four classes of policy problems that econometric analysis must address.

P1. Evaluating the impacts of implemented interventions on outcomes in a given environment, including their impacts in terms of the well-being of the treated and society at large.
  - Ex post analysis.
  - Evaluate the impact of increasing the level of childcare subsidy on childcare usage, and welfare gains, after the policy change has been implemented.
  - Partially addressed by RCM, and SCM. They cannot address welfare effects. Fully addressed by EMC.

The four classes of policy problems

P2. Understanding the mechanisms producing treatment effects and policy outcomes.

- What are the causes of the effects we observe?
- Why do some families seem to be insensitive to changes in childcare subsidy?
- Partially addressed by RCM, SCM (using mediators). Fully addressed by EMC.

The four classes of policy problems

P3. Forecasting the impacts (constructing counterfactual states) of interventions implemented under one environment when the intervention is applied to other environments, including their impacts in terms of well-being.

- Addresses external validity.
- What are the impacts of increasing the subsidy in an environment without activity test, when we only have information on impacts in an environment with activity test?
- Depending on method used, partially addressed by RCM, SCM. Fully addressed by EMC.
- Replaces meta-analysis with explanations based on economic principles.

The four classes of policy problems

P4. Forecasting the impacts of interventions (constructing counterfactual states associated with interventions) never previously implemented to various environments, including their impacts in terms of well-being.
- The problem policy analysts try to solve frequently.
- What are the impacts of removing the activity test in terms of childcare usage, costs, and welfare?
- Cannot be done without models, EMC.

Final remarks

- The main advantage of the EMC is that it can be linked to economic models.
- It is through those models that it is able to address all four classes of policy problems.
- In the example above, we could have simply estimated a multinomial choice model directly, without having to specify the utility function or budget constraint.
- However, that approach would not give us mechanisms of effects, nor external validity, nor welfare effects.

Final final remarks

- Each of the three frameworks of causality has advantages and disadvantages.
- Be familiar with all three.
- Use the right tool for the job.
- Sense-check your results, and don't take things too seriously.

Final final remarks

- Each of the three frameworks of causality has advantages and disadvantages.
- Be familiar with all three.
- Use the right tool for the job.
- Sense-check your results, and don't take things too seriously.
- Remember:

  "An economist is someone who goes into a dark room to find a black cat that doesn't exist. An econometrician is someone who claims to have found it – with 95% confidence."

Fin.