Causality & Correlation

Ethan Ligon

April 4, 2022

How can we draw inferences regarding causality?

We can infer joint probability distributions over observable variables, but by themselves these probability distributions can tell us *nothing* about cause.

No causes in. . .

It immediately follows that cause *cannot* be inferred simply by reference to data. As philosopher Cartwright (1994) puts it in her critique of Bayesian networks, "No causes in, no causes out." (This is a position perhaps first taken by David Hume, who framed "cause" as an idea.)

Ragnar Frisch (1930)

"The scientific answer to 'what is cause'. . . must read: it is such and such a way of thinking."

Identification Distinct from Causality

If data/probability distributions tell us nothing about causality absent a model, then it follows that the *analysis* of causality resides entirely in the specification and analysis of the model.

Identification

(Causal) Identification has to do with whether unspecified aspects of the model can be pinned down by trying to make model predictions match things observable in TRW. The issue of identification is conceptually distinct from the issue of causality.

Frisch (Again)

A well-specified model implies a joint probability distribution over some set of variables. Then identification depends on the "correspondence between the model world of probability and the real world of frequency…"

Notation for Causality

Structural approach in economics encodes causality in its syntax and in conventions for labeling variables; i.e., in a model written as

$$y = f(x, u)$$

a contemporary economist would probably read this as "y is caused by x and u", even though the mathematical equation itself carries no such causal implication.

Correlation does not imply causation

Consider the regression equation

$$y = a + bx + e,$$

with e *defined* by the condition that $\mathbb{E}(e|x) = 0$; for simplicity assume $\text{var}(x) = \text{var}(y)$.

Marginal Causal Effect?

It's tempting to define the 'marginal' causal effect of x on y as $\partial y/\partial x = b$, the least squares regression coefficient. But these are random variables! We can't just take derivatives of them like this!

Reverse Regression

Consider the 'reverse regression'

$$x = c + dy + e'$$

Given our assumptions about equal variances, estimating this yields $d = b$. So which way does the causation go?

Conditioning does not imply control

Consider the regression equation

$$y = a + bx + e,$$

with e *defined* by the condition that $\mathbb{E}(e|x) = 0$. Then we can define the conditional expectation $\mathbb{E}(y|x) = bx$.

Careful!

It's tempting to read this regression causally, inferring that a one unit increase in x implies a b unit increase in y. But we're conditioning on a random variable! We can't just go adding numbers to random variables. Even if we can change the 'location' of the *marginal* distribution of x what effect does this have on the joint distribution of (x, y)?

Structural Models (Cowles Commission)

A structural model can be expressed as a triple (U, V, F), where:

- U is a set of *exogenous* variables;
- V a set of *endogenous* variables; and
- F is a set of functions such that $V_i = F_i(U, V_{-i})$.

Causality in Structural Models

Note that this is entirely a property of the model! Frisch (1930) observes that "the main aspect of of the problem of scientific causality is the *direction* of causality."

Direct cause

A variable $X \in U \cup (V \setminus Y)$ is a *direct cause* of Y if it appears in the function describing Y, i.e., $Y = F_Y(X, \cdots)$.

Cause

A variable $X \in U \cup (V \setminus Y)$ is a *cause* of Y if it directly causes either Y or any other cause of Y.

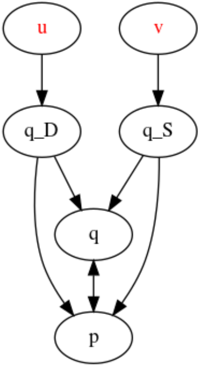
Graphical Models (Pearl)

Every structural model is associated with a directed graph, which summarizes causal relationships.

- A directed graph is is comprised of *nodes* and *directed edges*.
- The graph corresponding to a structural model (U, V, F) will have a node for every variable in $U \cup V$.
- If, in the structural model, a variable X directly causes a variable Y then there will an edge directed from X to Y.

The Canonical Model of Economics: Demand & Supply

An example of a *structural* model of demand & supply. A market equilibrium exists if there exists a pair (p^*, q^*) such that



$$\left.\begin{aligned} q_D &= q_D(p, u) \\ q_S &= q_S(p, v) \end{aligned}\right\}$$

and markets clear,

$$q_D(p^*(u, v), u) = q_S(p^*(u, v), v) = q^*(u, v).$$

Model is closed (or "completely specified") if we specify the functions (q_D, q_S) and the distributions of u and v, say F_u and F_v.

Some Questions Answered by the Canonical Model

Control What is the expected demand if we *set* the price $p = p_0$?

Observe What is the expected demand if we *observe* $p = p_0$?

Counterfactual If prices and quantities are observed to be (p_0, q_0), what would demand be if we were to *change* the price to p_1, *ceteris paribus*?

NB: "Observed" implicates identification

Controlling Price

Expected demand if we *set* price to p_0, no longer treating price as a random variable. Note that *setting* the price has no effect on the distribution of u, so:

$$\mathbb{E}(q_D(p_0)) = \int q_D(p_0, u) dF_u(u).$$

Observing Price

Expected demand if we *observe* a realization of $p = p_0$:

$$\mathbb{E}(q^*|p = p_0) = \\ \mathbb{E}\left[q^*(u, v)|q_D(p_0, u) = q_S(p_0, v)\right].$$

Given that we observe, say, a high price p_0 the *conditional* expected value of u will also be higher, and the conditional value of v will be lower (if q_D and q_S are monotonically increasing in u and v, respectively).

Price Change, *Ceteris Paribus*

If (p^*, q^*) are *observed* to be (p_0, q_0), what would demand be if we were to change the price to p_1, *ceteris paribus*?

- If we observe $(p^*, q^*) = (p_0, q_0)$ we can maybe infer (u_0, v_0) (using, e.g., monotonicity). Nothing's random anymore!

Then *ceteris paribus* we have the change

$$\Delta = q_D(p_1, u_0) - q_D(p_0, u_0).$$

Note that answer here doesn't depend on supply at all, except to infer u_0.