Causality in econometric pedagogy, 1953–2017

Chris Auld
University of Victoria

West Coast Experiments meetings, UCLA 2017

Overview.

- How do econometrics textbooks discuss causality? An analysis of 32 texts from 1953 to 2017.
- How have discussions of causality changed over time?
- Context and the *implicit* presentation of causal concepts.

Context.

- Critiques of econometric textbooks:
  - Chen and Pearl (2013): avoidance of and confusion over causation
  - Angrist and Pischke (2017): too much emphasis on GLS, not enough on modern methods for quasi-experiments
- Scant literature documenting econometric pedagogy (only Duo Qin?)

Preliminaries.

- To fix ideas, consider the univariate regression

$$y = \beta x + u \tag{1}$$

- If we *define* u as $[y - E(y|x)]$ in "agnostic regression," then

$$E[y|x] \equiv \beta x. \tag{2}$$

- But then β cannot generally be interpreted as a causal parameter.

Preliminaries cont.

- If we interpret u as causes of y other than x in a "structural" interpretation, then

$$E[y|x] \equiv \beta x + E[u|x], \tag{3}$$

  the last term does not generally vanish and estimates of $E[y|x]$ do not generally recover β, which is now the ceteris paribus (causal) effect of x on y.
- Basic goal here: how have textbooks presented discussion of these issues?

Background: a very very brief history of relevant econometric thought

1699	Davenant, first demand curve estimate
1943	Haavelmo, formalizes simultaneity bias
1950	Anderson and Rubin, 2SLS
1969	Granger, time series “causal” concept
1976	Lucas critique
1985	Heckman and Robb, treatment effects
1987	Rust, estimable structural models
1990	Card, Mariel boatlift: natural experiment, diff-in-diff
1994	Imbens and Angrist, LATE
1999	Angrist and Lavy, RD design

Cowles Commission almost–causal econometric jargon

- Split variables into two types and specify part or all of a system of simultaneous equations:

Endogenous. $\text{Cov}(x, u) \neq 0$. A regression of y on this sort of x doesn't recover a causal effect.

Exogengous. $\text{Cov}(x, u) = 0$. A regression of y on this variable does recover a causal effect.

- But these concepts are not identical to defining causality, e.g., if x is measured with error, we estimate β with bias even when estimates would not admit causal interpretation without measurement error.

A generic econometrics textbook outline:

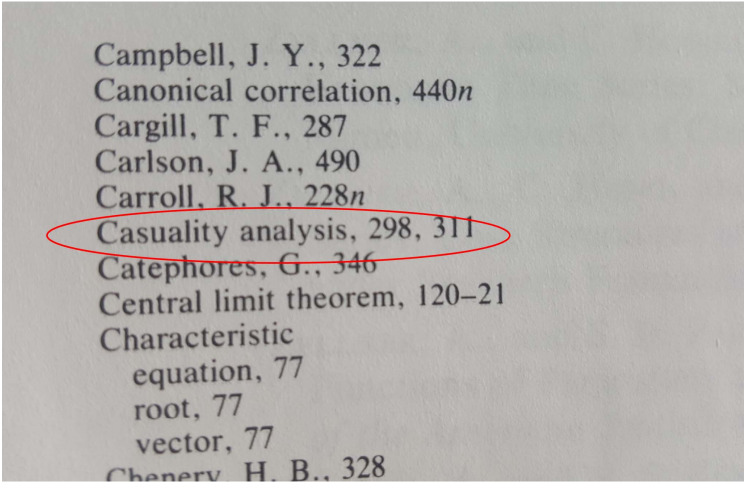
1. Start with $y = \beta x + u$. Assume u is i.i.d. and independent of x. Prove the OLS estimate converges to the value of β in the GDP.
2. Permit $V(u) = \Omega$, which is not proportional to I_n. Discuss a number of GLS issues.
3. Add a second equation, $x = \gamma y + \epsilon$ and proceed to discussion Cowles Commission approach to SEM.
4. Optional tangents: time series issues, forecasting, Bayesian methods, limited dependent variables, etc.

A dataset on econometrics textbooks.

- collected data on 32 (because $30 \approx \infty$) textbooks published between 1953 and 2017.
- convenience sample including most major texts
- Superset of Chen and Pearl (2013) and (but one) of Angrist and Pischke (2017)
- undergrad and grad
- some "applied" but no specialist texts (e.g., exclude books exclusively dealing with microeconometrics or time series)
- only one edition of each book

Stuff to collect.

- pages devoted causal methods: SEM, IV, panel methods, treatment effects
- Does the text contain any explicit discussion of causality?
- ... other than Granger "causality"?
- Is there a (faulty?) discussion of β?
- Is there a (faulty?) discussion of the error term?
- Are supply and demand or Keynesian MPC used as motivating examples?



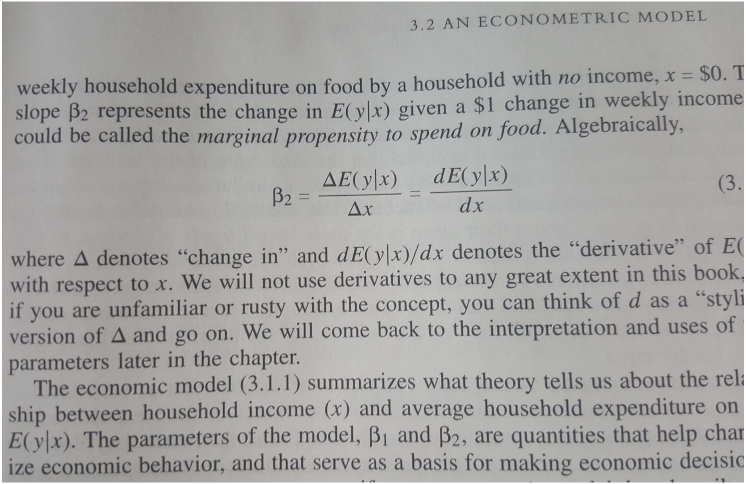
Problematic definitions of β or the error term

- Chen and Pearl (2013) note several current texts *define*

$$\beta = \frac{\partial E[y|x]}{\partial x}$$

  which is not generally true except in agnostic regression, and becomes rapidly confusing when these texts go on to discuss endogeneity.
- Similarly, *defining* $u \equiv y - E[y|X]$ is problematic.
- Look for this problem, not including texts which first *explicitly assume* u is uncorrelated with x.

Figure: Example of a book defining β as the gradient of $E[y|x]$ (Hill, Griffiths, and Judge 2001)



3.2 AN ECONOMETRIC MODEL

weekly household expenditure on food by a household with *no* income, $x = \$0$. T slope β_2 represents the change in $E(y|x)$ given a $1 change in weekly income could be called the *marginal propensity to spend on food*. Algebraically,

$$\beta_2 = \frac{\Delta E(y|x)}{\Delta x} = \frac{dE(y|x)}{dx} \quad (3.$$

where Δ denotes "change in" and $dE(y|x)/dx$ denotes the "derivative" of $E($ with respect to x. We will not use derivatives to any great extent in this book, if you are unfamiliar or rusty with the concept, you can think of d as a "styli version of Δ and go on. We will come back to the interpretation and uses of parameters later in the chapter.

The economic model (3.1.1) summarizes what theory tells us about the rel ship between household income (x) and average household expenditure on $E(y|x)$. The parameters of the model, β_1 and β_2, are quantities that help char ize economic behavior, and that serve as a basis for making economic decisio

Figure: Example: “foolproof” interpetation of OLS estimates (Dougherty 2007)

BOX 1.1 Interpretation of a linear regression equation

This is a foolproof way of interpreting the coefficients of a linear regression

$$\hat{Y}_i = b_1 + b_2 X_i$$

when Y and X are variables with straightforward natural units (not logarithms or other functions).

The first step is to say that a one-unit increase in X (measured in units of X) will cause a b_2 unit increase in Y (measured in units of Y). The second step is to check to see what the units of X and Y actually are, and to replace the word ‘unit’ with the actual unit of measurement. The third step is to see whether the result could be expressed in a better way, without altering its substance.

The constant, b_1, gives the predicted value of Y (in units of Y) for X equal to 0. It may or may not have a plausible meaning, depending on the context.

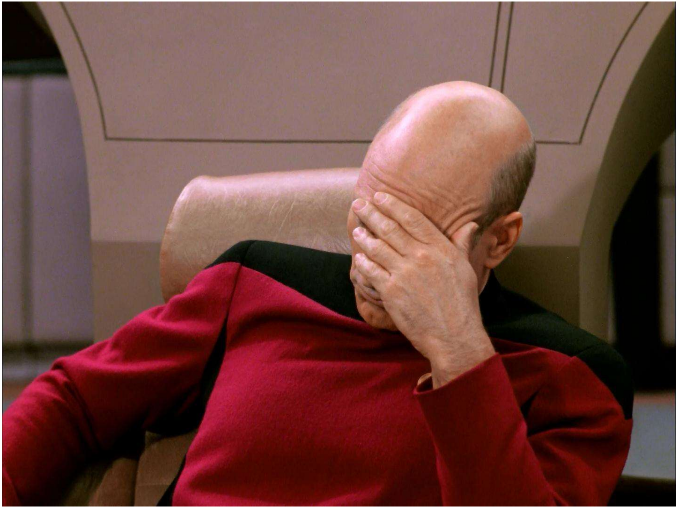
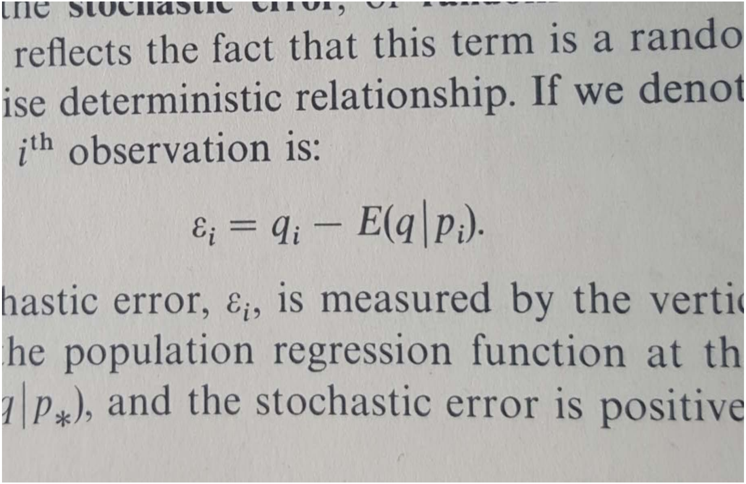


Figure: Example of an incorrect definition of the error term (Lardaro 1993)


reflects the fact that this term is a rando
ise deterministic relationship. If we denot
i^{th} observation is:

$$\varepsilon_i = q_i - E(q|p_i).$$

hastic error, ε_i, is measured by the verti
he population regression function at th
$q|p_*$), and the stochastic error is positive

Summary stats and linear trends. ($\hat{\beta}$ from regression of var on year/10)

variable	mean	$\hat{\beta}$	t-stat
pages	641.094	5.550	2.288
causal	0.438	0.003	0.577
"causal" defined	0.156	0.005	1.220
causal pages	84.094	0.394	0.686
proportion causal pages	0.134	-0.002	-2.162
Granger only	0.406	0.005	0.957
error defined	0.500	0.009	1.513
β discussion	0.313	0.013	2.596
β defined wrong	0.200	0.010	1.935
error defined wrong	0.133	-0.001	-0.186
S-D example	0.813	0.000	0.091
MPC example	0.938	0.007	0.368

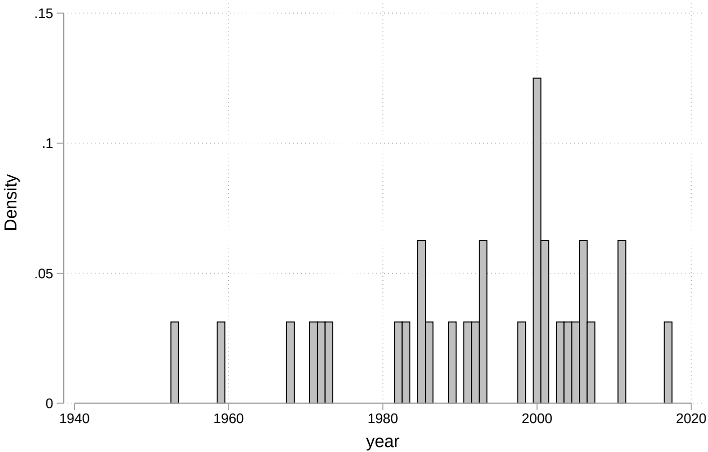


Figure: Distribution of publication dates (edition used, not nec. first edition).

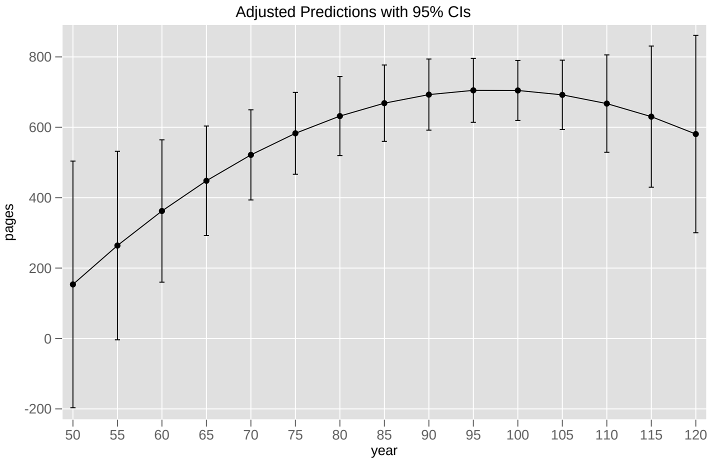


Figure: Pages.

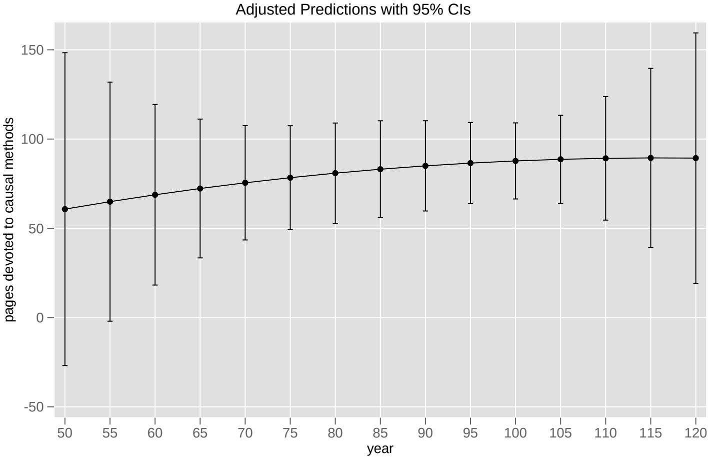


Figure: Pages devoted to causal methods.

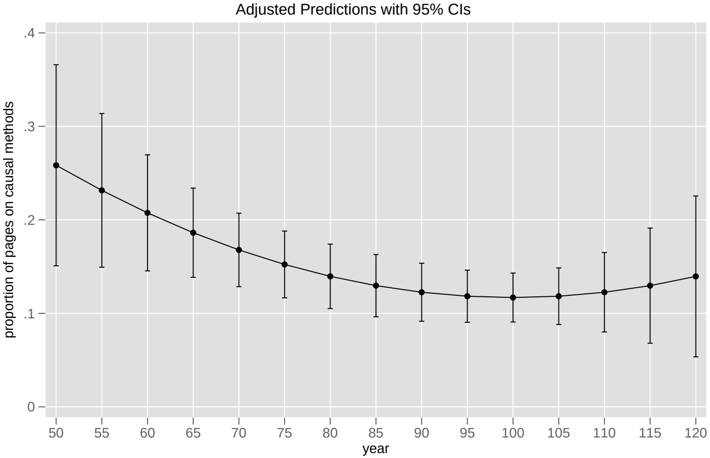


Figure: Proportion of pages devoted to causal methods

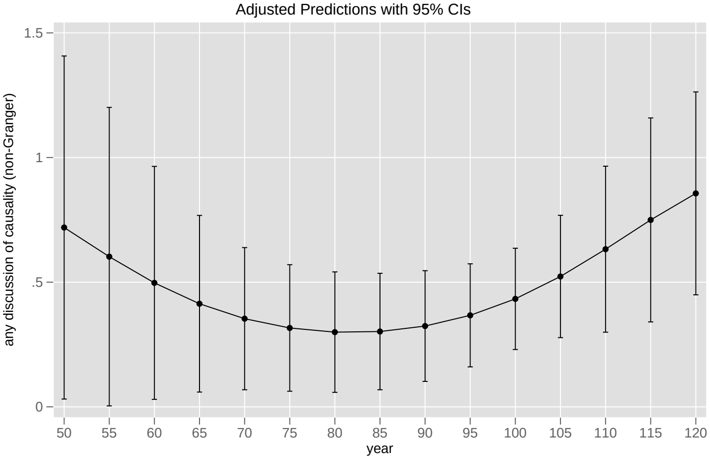


Figure: Is there *any* discussion of causality?

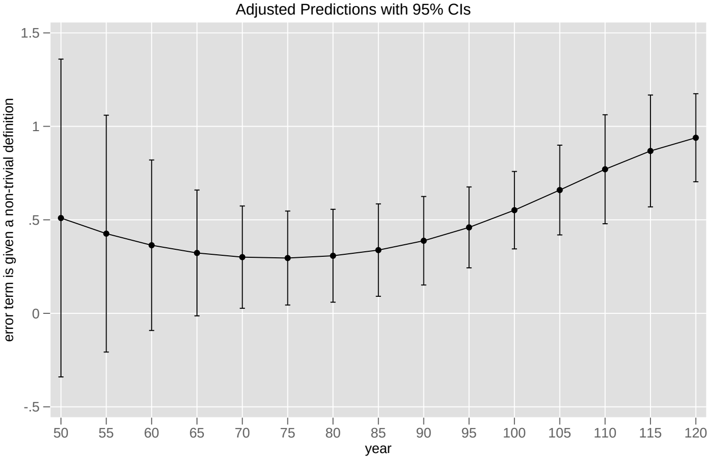


Figure: Is the meaning of the error term discussed?

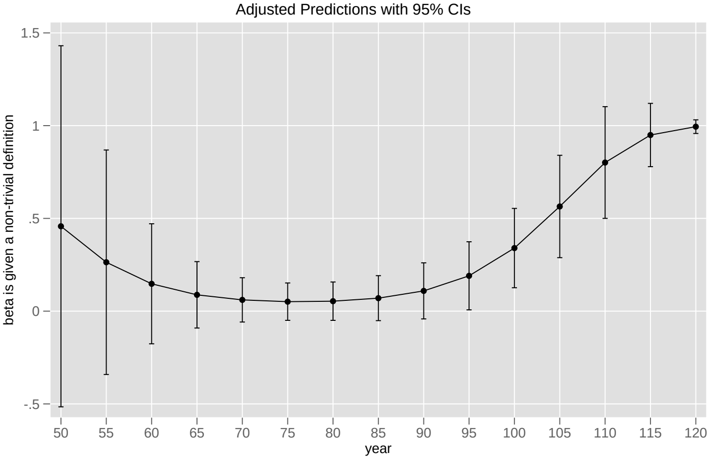


Figure: Is the meaning of β discussed?

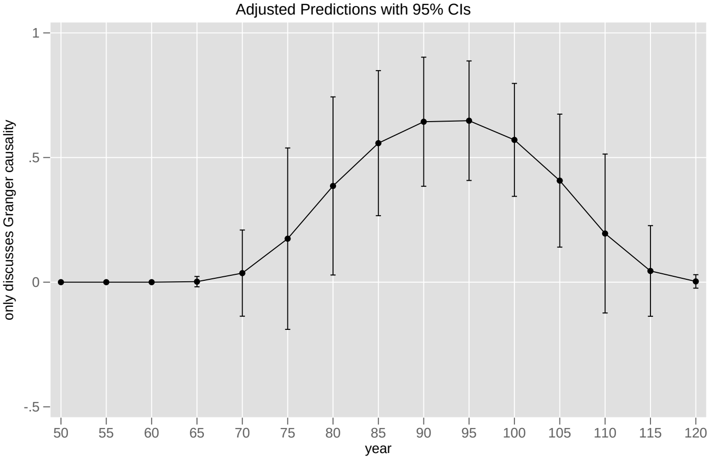


Figure: Is causality defined only in Granger's sense?

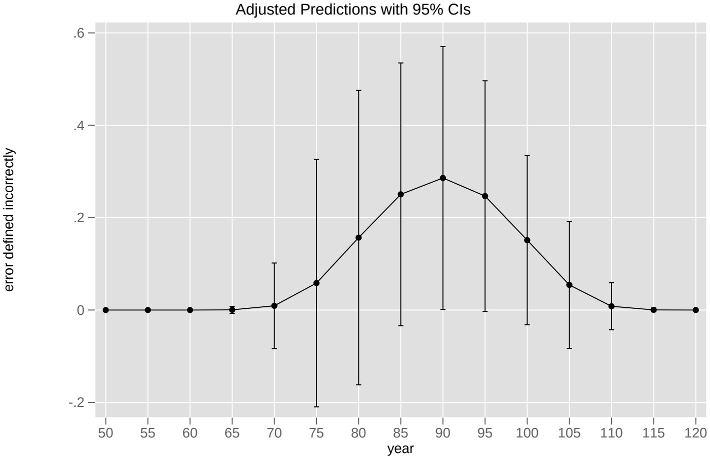


Figure: Is β incorrectly defined as the gradient of $E[y|x]$?

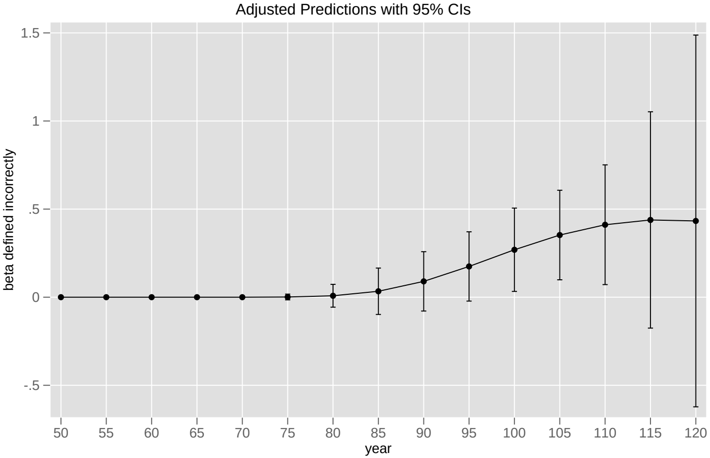
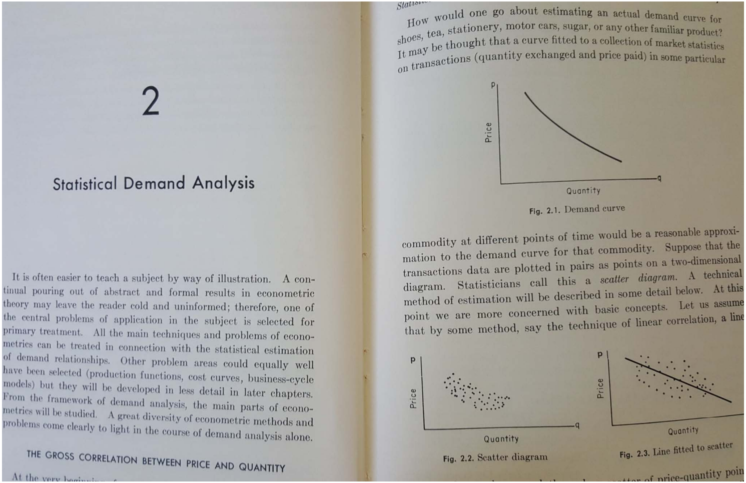


Figure: Is the error term defined incorrectly?

Figure: From extended discussion opening Klein 1953



2

Statistical Demand Analysis

It is often easier to teach a subject by way of illustration. A continual pouring out of abstract and formal results in econometric theory may leave the reader cold and uninformed; therefore, one of the central problems of application in the subject is selected for primary treatment. All the main techniques and problems of econometrics can be treated in connection with the statistical estimation of demand relationships. Other problem areas could equally well have been selected (production functions, cost curves, business-cycle models) but they will be developed in less detail in later chapters. From the framework of demand analysis, the main parts of econometrics will be studied. A great diversity of econometric methods and problems come clearly to light in the course of demand analysis alone.

THE GROSS CORRELATION BETWEEN PRICE AND QUANTITY

How would one go about estimating an actual demand curve for shoes, tea, stationery, motor cars, sugar, or any other familiar product? It may be thought that a curve fitted to a collection of market statistics on transactions (quantity exchanged and price paid) in some particular

Fig. 2.1. Demand curve

commodity at different points of time would be a reasonable approximation to the demand curve for that commodity. Suppose that the transactions data are plotted in pairs as points on a two-dimensional diagram. Statisticians call this a *scatter diagram*. A technical method of estimation will be described in some detail below. At this point we are more concerned with basic concepts. Let us assume that by some method, say the technique of linear correlation, a line

Fig. 2.2. Scatter diagram

Fig. 2.3. Line fitted to scatter

determine the three economic variables q_t^d, q_t^s, and p_t when *given* the random disturbances u_t, v_t, and w_t and the external variable r_t. We shall call the economic variables *endogenous* variables and the external variables *exogenous* variables. The laws of nature (meteorology in this instance) determine the values taken on at each point of time by r_t, independent of economic decisions or behavior in the supply-demand market. Rainfall affects the economy but is not affected by the economy. We cannot say the same of the endogenous variables.[5]

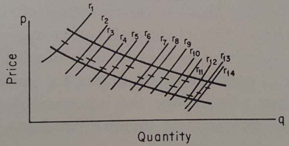


Fig. 2.6. Scatter of supply with rainfall induced shifts and stable demand

Regardless of the relative variabilities of u_t and v_t, the supply function drawn with respect to quantity and price axes will shift according to the different values assumed by r_t. This will help us to

Figure: By 1985, this is how a popular textbook opens

be considered throughout the chapter.

1.1.1 Introduction

Consider a sequence of K random variables $(y_t, x_{2t}, x_{3t}, \ldots, x_{Kt})$, $t = 1, 2, \ldots, T$. Define a T-vector $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, a $(K-1)$-vector $\mathbf{x}_t^* = (x_{2t}, x_{3t}, \ldots, x_{Kt})'$, and a $[(K-1) \times T]$-vector $\mathbf{x}^* = (\mathbf{x}_1^{*\prime}, \mathbf{x}_2^{*\prime}, \ldots, \mathbf{x}_T^{*\prime})'$. Suppose for the sake of exposition that the joint density of the variables is given by $f(\mathbf{y}, \mathbf{x}^*, \theta)$, where θ is a vector of unknown parameters. We are concerned with inference about the parameter vector θ on the basis of the observed vectors $\mathbf{y}$ and $\mathbf{x}^*$.

In econometrics we are often interested in the conditional distribution of one set of random variables given another set of random variables; for example, the conditional distribution of consumption given income and the conditional distribution of quantities demanded given prices. Suppose we want to know the conditional distribution of $\mathbf{y}$ given $\mathbf{x}^*$. We can write the joint density as the product of the conditional density and the marginal density as in

$$f(\mathbf{y}, \mathbf{x}^*, \theta) = f(\mathbf{y}|\mathbf{x}^*, \theta_1) f(\mathbf{x}^*, \theta_2). \tag{1.1.1}$$

Causality without causal language

- Econometrics textbooks past and present tend to treat inference as a purely statistical problem, but the model as an object (often) from economic theory.
- Explicit causal language is shunned!

> *"Econometricians as a rule avoid the concept of causality altogether."*
>
> –Jan Kmenta, 1988

But! Context.

- Yet every econometrician, now and historically, would of course nod in response to the claim "correlation does not imply causation."
- Papers routinely attacked for "endogeneity problems," which usually means the critic doesn't find an attempt at causal inference compelling.
- How do we reconcile the acute focus on causal issues with the dearth of causal pedagogy?
- We can make sense of this, if it's true, by acknowledging that econometrics courses are taught alongside economic theory courses.

Supply and demand.

- The first semester of Econ 101 focuses on this model:

$$\text{demand}: \; Q^D = \alpha^D - \gamma^D P \tag{4}$$

$$\text{supply}: \; Q^S = \alpha^S - \gamma^S P \tag{5}$$

$$\text{equilibrium}: \; Q^D = Q^S \tag{6}$$

- The γ are *causal parameters* denoting the responses consumers and firms would make to changes in price if price were set as if by a RCT.
- Most (80%) econometrics textbooks use exactly this model to illustrate the result that correlation does not imply causation (interpreting the α as random variables).

Implicit causal language: Example.

- Demand: $Q^D = \alpha_D - \gamma_D P$.
- Economic theory tells us: if price is set as if by a RCT, the causal effect of a one-unit increase in price on quantity demanded is γ_D.
- Textbooks, *without explicit causal language*, prove OLS does not consistently estimate γ_D.
- What can this mean other than: 'the OLS estimate does not recover the *causal effect* of price on quantity demanded'?

Implicit causal language: examples.

This problem introduces a basic econometric concept—the concept of identification. The earliest empirical work in econometrics, the estimation of statistical demand curves, uncovered the problem that observed correlations between prices and quantities of commodities did not necessarily reveal the nature of the demand curve.

Klein, 1953, p17

Implicit causal language: examples.

As is true of a great many econometric models, the parameters in [Keynesian MPC model] can be seen to have a direct interpretation in terms of economic theory.

Davidson and MacKinnon 2004, p2

Costs and benefits of implicit causal jargon

- Characterizing the *statistical* assumptions required for an estimator to converge to a parameter which is to be interpreted as a causal effect places the issues squarely in terms of the data.
- e.g., using w as an instrument, we show precisely the conditions under which

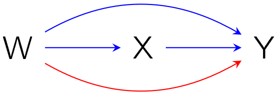
$$\tilde{\beta}^{IV} \rightarrow \frac{Cov(w, y)}{Cov(w, x)} = \beta \tag{7}$$

Costs and benefits of implicit causal jargon

- But: a complete characterization of the statistical requirements for an estimator to recover a causal effect may obscure exactly what causal relations we require to exist or not exist.

Example.

Figure: "W causes Y only through its effect on X" — W is valid if the red arrow doesn't exist:



Example.

Figure: "W causes Y only through its effect on X" — W is valid if the red arrow doesn't exist:

$$\begin{tikzcd} W \arrow[r] \arrow[rr, bend left] \arrow[rr, bend right] & X \arrow[r] & Y \end{tikzcd}$$

Figure: But this obscures fact that W is also invalid if W and Y have a common cause (e.g., Pearl 2000, Margolis 2015):

$$\begin{tikzcd} W \arrow[r] \arrow[rr, bend left] \arrow[rr, bend right, dashed, leftrightarrow] & X \arrow[r] & Y \end{tikzcd}$$

Caveat.

- Prediction and forecasting are also important in some parts of econometrics, notably macroeconometrics.

  > *The AP approach to [prediction] is to ignore it, in a thinly-veiled attempt to equate econometrics exclusively with [causal inference]. Sorry guys, but no one's buying it. That's why the textbooks continue to feature [prediction] tools and techniques so prominently, as well they should.*

  Diebold blog, Angrist and Pischke are at it again, Feb 2017
- Yet everyone should agree that causal inference should be (1) correctly discussed and (2) differentiated from predictive or forecasting exercises.

Conclusions.

- Explicit discussion of causality waned in the 1970s through to about 2000, then experienced a resurgence.
- Confusion or incorrect assertions about the nature of β or the error term are quite common.
- Textbooks have consistently implicitly discussed causal notions, and must be placed in the context of the broader economics curriculum.
- Hopefully the current resurgence of causal discussion will continue into the future.