Models, Measurement, and the Language of Empirical Economics

Phil Haile

November 2024

Overview

The language of empirical economics is often confusing and inconsistent.

Today I hope to

1. offer a logically consistent taxonomy of empirical work
2. remind you of some definitions
3. discuss some commonly abused terms in empirical economics
4. discuss some benefits of economic models for empirical work.

Types of Empirical Work

All empirical work is either descriptive or structural.

Descriptive

Descriptive empirical work aims to characterize relationships between observables

- i.e., establish facts about the data, e.g.,
  - college grads earn 98% more per hour than others
  - income inequality higher now than 30 years ago
  - health care costs growing more slowly after ACA
  - airline prices higher now than before merger wave

  (such facts sometimes suggest interpretations).

Structural

Structural empirical work aims to quantify certain features of an assumed data generating process (i.e., a "structure") that are useful for interpreting data; e.g.,

- estimate demand for schools→predict outcomes under a voucher system
- estimate model of schooling, marriage, and labor supply choices→ quantify specific notions of the male-female wage gap
- estimate demand and firm costs→predict effects of a merger
- estimate a model of treatment takeup and outcomes to assess the effect of the treatment on certain groups.

Structure and Models

Examples of a structure:

- $Y = X\beta_0 + \epsilon$ with particular β_0 and distn of $\epsilon|X$
- equilibrium model, with particular parameter values and distributions of unobservables

A model is a collection of admissible structures, e.g.,

- $Y = X\beta + \epsilon$ with $\epsilon \perp X$ and unknown β, or
- equilibrium model of demand and supply with unknown parameters θ and certain exogeneity conditions

Thus, a model is a set of maintained hypotheses about the true structure (true DGP). We'll be a little more formal about this later.

Counterfactuals

The term counterfactual is central to much of empirical economics, and to most efforts to interpret data rather than merely describe it.

A counterfactual question takes the form

"what would happen to Y if X (and nothing else) were altered?"

- almost always, this is a question about a state of the world not directly observed—a counterfactual world
- this form of question nests many things

Most questions in empirical economics take this form.

Examples of Empirical Questions in Economics

- how much do prices change with the number of competitors?
- how much would prices rise if the sales tax were raised by 1%?
- how has productivity (e.g., TFP) in the U.S. auto industry changed over the last 30 years?
- what is the average effect of college attendance on wage?
- what is the elasticity of demand for health insurance in MA?

Counterfactual Questions in Economics

The above are all counterfactual questions:

- *effect of entry (or sales tax) on prices:* we can't just compare local markets with differing numbers of competitors (different sales taxes)
- *productivity*: we can't just look at inputs and outputs to learn the production function (and thus TFP), since unobservables affect both the input levels chosen by firms and their realized output; measuring TFP changes requires counterfactually holding those fixed
- *returns to college*: we can't just compare wages for those who attended and those who did not
- *elasticity of demand*: we can't just compare quantities purchased for goods with high prices vs. those with low prices.

Causal Effects and Counterfactuals

A causal effect is the response of an outcome Y to a an intervention that altered X (but nothing else).

This is what we used to call "the effect of X on Y" until rampant abuse of "effect" created a need for the redundant label "causal."

Such an effect reflects the concept of a "ceteris paribus" change, central in economics at least since Alfred Marshall.

Causality is a concept defined by a question of the form "what would happen if certain things were changed while others were held fixed?" This is typically a question about a counterfactual world.

Causal Effects and Structural Estimation

Outside economics, some disciplines lacking the tradition of formal frameworks for defining counterfactual questions have invented new ones: e.g., the PO framework/"Rubin causal model" in statistics or the DAG/SCM models of Pearl in CS.

This is fine, and sometimes valuable:

- a model of some kind is necessary to define (and then explore) counterfactual quantities of interest;
- one can have preferences over models (e.g., depending on the context)

But regardless of the model chosen, causal identification/estimation/inference is always a special case of structural identification/estimation/inference.

What about Program Evaluation?

Program evaluation is a term used to describe measurement of an effect of some treatment on some outcome of interest.

Arguably this covers all counterfactual questions, but typically applied more narrowly: either to counterfactual questions defining certain standard notions of "treatment effects" like ATE, LATE, etc. or to empirical work using certain tools (DiD, RD, etc.).

Program Evaluation $\subset$ Structural Estimation

Program evaluation is always a form of structural estimation. It requires a set of maintained hypotheses about the world (i.e., a model) allowing one to define and then identify a counterfactual quantity of interest.

For example, TT, ATE, LATE, QTE, etc. are just examples of what we mean by "features of the data generating process (true structure)" when defining structural estimation. They can be precisely defined only with a model of how the data are assumed to be generated.

Any suggestion that these objects are "model free" is simply nonsense (e.g., Rubin Causal Model is a model!). One also must use the model to determine which statistical procedures measure which notions of treatment effect.

Randomized Controlled Trials

Typically one uses a model to define what it means to "hold all else fixed" in the counterfactual of interest—e.g., to define the necessary controls and latent factors whose impacts must be accounted for.

In a RCT, after defining the counterfactual question of interest, one aims to directly create the counterfactual world (at least up to latent factors that the model says will average out). For example, on may assign drugs randomly to patients instead of through the mix of decisions that would otherwise prevail. But only after properly defining the question, the relevant latent factors, and their roles, can one ask how to design the RCT to answer it.

Still, there is indeed something special about a RCT.

A Special Case: An Ideal RCT

Consider a RCT with interest only in the average effect of the treatment in the population studied.

In this case, the empirical work answering the question is structural: by design, it quantifies features of an underlying functional or probabilistic relationship between treatment, controls, unobservables, and outcomes that is assumed to be invariant to the intervention.

But it is also descriptive: at the end of the day, it involves only a characterization of relationships between observables.

Program Evaluation is Structural Estimation

Sometimes (e.g., in RCT) data description directly reveals a causal/structural/countefactual quantity like TT or ATE.

But that is a result—one that follows only from a set of assumptions on the underlying structure that allows us to distinguish the notion of interest from others and to logically conclude that the data directly reveal this quantity.

Typically program evaluation requires more than descriptive analysis: one must counterfactually hold all else equal to learn the true effect(s) of a treatment D on an outcome Y, given some controls X. This is only possible after specifying enough of the underlying structure that one can properly define the particular treatment effect of interest and determine what maintained assumptions and empirical techniques can reveal it.

Quasi-Experimental

Quasi-experimental is a label given to certain types of variation in the data that one may exploit to obtain identification of the structural feature(s) of interest.

A leading example is variation through an instrumental variable.

Quasi-experimental variation is something that can allow one to do structural estimation in a way that is more credible (i.e., relying on assumptions we are more comfortable accepting when interpreting data).

Any suggestion that "quasi-experimental" describes an *alternative* to structural estimation rather than a special case is confusion.

A Final Observation about Counterfactuals

A counterfactual may involve a feasible policy change (e.g., raise the minimum wage). But it could instead involve a purely hypothetical intervention (e.g., change the race (only) of some people to see the resulting wage change).

As always, one must think carefully about the counterfactual of interest: e.g., race switched at age 25? at birth? But those suggesting that one cannot define the (causal) effect of being black rather than white are confused. A "hypothetical policy intervention" is often a good way to think about a counterfactual, but not always. It is an illustration of the concept, not its definition!

Identification

The concept of identification is central to empirical economics, precisely because the questions of interest involve counterfactuals.

Identification is discussed much more in modern empirical work than in the past. This is good: it reflects greater attention to arguably the fundamental challenge of empirical economics: whether/how the things we observe are capable of revealing the answers to the questions we care about.

But with this attention has come some sloppiness about what identification means.

Identification (Review)

(see e.g., Hurwicz (1950), Koopmans and Reiersol (1950), Berry and Haile (2018))

- a structure S is a data generating process, i.e., a set of probabilistic or functional relationships between the observable and latent variables that implies ("generates") a joint distribution of the observables
- let $\mathfrak{S}$ = the set of all structures; $S_0 \in \mathfrak{S}$ the true structure
- a hypothesis is any nonempty subset of $\mathfrak{S}$ (a model is one important example)
- hypothesis $\mathcal{H}$ is true (satisfied) if $S_0 \in \mathcal{H}$
- a structural feature $\theta(S_0)$ is a functional of the true structure

Definition. A structural feature $\theta(S_0)$ is identified (or *point identified*, or *identifiable*) under the hypothesis $\mathcal{H}$ if $\theta(S_0)$ is uniquely determined within the set $\{\theta(S) : S \in \mathcal{H}\}$ by the joint distribution of observables.

Identification Presumes Structure

Identification cannot even be defined without the notion of a true structure within a class defined by maintained hypotheses (what we usually call a "model") and a structural feature of interest.

The model may be simple or complicated, may involve economics or only hypothesized probabilistic relationships (e.g., Rubin causal model).

But the very concept of identification presumes that there are structural features one wishes to uncover.

What about "Reduced Form"?

The term "reduced form" is widely used—most often in a way that I believe is impossible to rationalize with a coherent definition.

Most of what is labeled "reduced form" empirical work is in fact structural. That which isn't is descriptive.

There are layers of possible confusion here.

Reduced Form: Econometrics Definition

Definition. A **reduced form** is a functional or stochastic mapping for which the inputs are (i) *exogenous* variables and (ii) unobservables ("structural errors"), and for which the outputs are *endogenous* variables. e.g., $Y = f(X, Z, U)$.

This is the textbook definition and the only formal definition in econometrics that I am aware of. It goes back at least to 1950.

Where does it come from?

A reduced form is obtained by solving the relevant structural model for each endogenous variable as a function of the exogenous observables and structural errors.

The classic example is perfectly competitive supply and demand in a one-good economy:

$$\begin{aligned} Q &= D(P, X, U_d) \qquad \text{(demand)} \\ P &= MC(Q, Z, U_s) \qquad \text{(supply)} \end{aligned}$$

Solving for equilibrium yields the reduced form relations

$$\begin{aligned} P &= p(Z, X, U_s, U_d) \\ Q &= q(Z, X, U_s, U_d). \end{aligned}$$

Reduced Form: One Small Source of Confusion

Solving for the reduced form isn't actually essential. We just need to know what goes on which side of the equation(s) and what restrictions, if any, must be imposed.

This—especially the latter—might be difficult without writing down a structural model. But in principle one can hypothesize a reduced form rather than hypothesizing structural relationships and deriving the reduced form.

Another Nuance

Definition. A **reduced form** is a functional or stochastic mapping for which the inputs are (i) *exogenous* variables and (ii) unobservables ("structural errors"), and for which the outputs are *endogenous* variables. e.g., $Y = f(X, Z, U)$.

But what does *exogenous* mean?

- when talking about theory, an exogenous variable is one taken as given in (not determined within) the model
- when talking about econometrics, an exogenous variable is one satisfying some kind of independence condition with respect to unobservables.

These can be different: the fact that something is treated as "given" does not mean it satisfies any independence conditions!

Reduced Form in Econometrics: Some Observations

1. RF requires at least implicit reference to a (structural) model: one cannot know what the arguments of the RF are without at least having some notion of a structural model in mind
2. RF equation can involve a fully flexible function, an approximation (e.g., linear) to fully flexible function, or can take the form obtained by solving the (parametric or nonparametric) structural model
3. RF can be used for some types of counterfactuals!
   (those involving changes only in the values of the arguments of the RF)
4. sometimes there is no difference between the structural equation and reduced form: e.g., exogenous treatment with $Y = f(X, U)$ and scalar $U \perp\!\!\!\perp X$
5. the reduced form is not automatically identified (see supply and demand, e.g.. where we have multiple structural errors in each RF eqn).

Reduced Form: Theory Definition

For a theorist, a reduced form refers to a representation of a mapping from inputs to outputs without incorporating all underlying mechanisms or implied functional forms.

An example: variable profit of a price-setting oligopolist

- we could represent this as

$$max_p D(p; x, n)(p - c(w))$$

  where x are demand shifters, n is the (endogenous) number of competing firms, and w are cost shifters
- or, noting that we will have $p = p^*(x, w, n)$, for some purposes we could just specify a reduced form profit function $\pi(x, w, n)$.

Reduced Form: Theory Definition

Confusion and Abuse of Terminology 101

This notion—of a simplified relationship that does not fully represent all underlying mechanisms—is what some empirical folks will reference to explain what "reduced form" empirical work is.

But all models suppress underlying mechanisms!

So by this definition all models are reduced form, specifically

$$\{structural\ models\} \subset \{reduced\ form\ models\}$$

It could sometimes make sense to say that one model is "more reduced-form" than another. But as a binary classifier, this notion of "reduced form" has no content!

"Reduced Form"

Confusion and Abuse of Terminology 101

Note that the "reduced form" label is typically combined with use of IV, RD, DiD, or other "identification strategies" due to endogeneity/selection/latent factors etc.

For this to make sense, one must mean that the model estimated is "structural" in the econometrics sense but "reduced form" in the theory sense. This would be a "reduced form structural model."

This would be logically coherent, if a bit silly. It is as close as I think one can get to rationalizing the common use of "reduced form" in applied economics. But it does not seem consistent with efforts to categorize empirical work with labels "reduced form," "structural," "quasi-structural" etc.

“Reduced Form”

Confusion and Abuse of Terminology 102

“Reduced form” is sometimes used to mean “equation I won’t derive, justify, or take questions on, but which I will nonetheless treat as structural (i.e., ‘causal’) when I talk about conclusions.”

At best this is bad science. There is, of course, bad science of every flavor.

"Reduced Form"

Confusion and Abuse of Terminology 103

Sometimes reduced-form terminology used with correct intent but without enough care to be transparent about what is being assumed.

Suppose someone has in mind the supply and demand model

$$\begin{aligned} Q &= D(P, X, U_d) \qquad \text{(demand)} \\ P &= MC(Q, Z, U_s) \qquad \text{(supply)} \end{aligned}$$

but skips the structural model to posit a reduced-form (theory definition) pricing equation of the form:

$$P = g(Z, X, \epsilon) \qquad \epsilon \in \mathbb{R}$$

This looks reasonable but implicitly makes a strong functional form restriction: the structural errors (U_d, U_s) enter the eqm solution only through a scalar index $\epsilon(U_d, U_s)$.

An Aside: Name You Errors!

The supply and demand example illustrates a general substantive point: it is important to ask what the unobservable(s) are in the relevant model, rather than treating them as unnamed "residuals," "heterogeneity," or "error terms." By asking what ϵ above is, one would realize that it must reflect both demand shocks and cost shocks.

It is hard to speak coherently about the properties of unnamed "error terms"!

Reduced Form vs. Descriptive

Baby and Bathwater

“Reduced-form” sometimes used to mean descriptive, sometimes to mean that the structural model is viewed as simple (not representing all mechanisms), sometimes to mean “sloppy,” and sometimes in a way consistent with its econometrics definition.

Many interesting papers involve descriptive analysis that evaluates model predictions or suggests patterns/phenomena that one might investigate further using other methods. And for many questions, simple structural models make sense and allow one to answer questions of interest. Mis-labeling these things as “reduced-form” causes confusion at best.

Reduced Form: The Bottom Line

It does not seem possible to rationalize the most common uses of "reduced form" in empirical economics with a coherent definition beyond the nonsensical: "the set of approaches I have decided to call reduced form."

Much of the work labeled "reduced form"—including all of that involving causal effects—is unambiguously structural. That which is not is descriptive.

"Structural"

Confusion and Abuse of Terminology 201

Often "structural" misused...

- to describe *how* one estimates rather than *what* one estimates: "We *structurally estimate* a model of..."
  - this is nonsense (and same for "causally estimate/identify"!)
  - "we estimate a model of ..." will do
- that one is estimating the "deepest" primitives one thinks of
  - this defines terminology based on the speaker's knowledge/imagination
  - it is also just incorrect; it is the flip-side of labeling as reduced form anything that doesn't account for all mechanisms.

"Structural"

Confusion and Abuse of Terminology 202

Sometimes "structural" misused...

- to mean use of a complex model, or a model with many parametric/functional form assumptions
  - complexity/parameterization is completely orthogonal to the question of structural vs. descriptive vs. reduced form;
- to mean that all functional forms and distributions have been specified up to a finite parameter vector
  - this is wrong; e.g., there is a big literature on identification of various nonparametric structural models.

Structural: The Bottom Line

Although "structural" is well defined, its abuses are also common.

"Structural" nests all types of empirical work that go beyond data description.

One should debate the merits of alternative approaches to structural estimation—alternative models, identification strategies, and statistical procedures used to estimate the quantities (structural features) of interest. Recognize that this is the only real scientific debate and you will cut through a great deal of confusion that currently limits communication, cross-fertilization, and progress in economics.

Does this language really matter?

Some possible views:
(1) the meanings of words evolve, so what?
(2) a word's use in practice determines its meaning, not the other way around!

These views are appropriate if we are writing a dictionary, where the fact that certain words are used in ill-defined or contradictory ways is a fact to be reported.

But the language we choose to use when communicating about science is another matter. I don't think one can defend choosing fuzzy language about any element of research that matters.

The Words We Choose

(1) Misuse/inconsistent use of terminology destroys information. This may be mildly annoying or amusing in everyday conversation

> *"I literally died laughing when I heard what Josh Angrist said about empirical IO!"*

but is sloppiness that should be avoided in serious scholarship, at least when it comes to notions meant to convey something of importance.

(2) The language we use shapes the way we think and how we are understood. Precise language encourages precise thinking and transparent exchange of ideas. This is part of our job.

(3) We should debate modeling approaches and tradeoffs between the assumptions relied upon and the questions answered. This is what the "structural vs. reduced form" divide is actually about—i.e., preferences over different types of structural models. Misuses of language creates a false barrier that shuts down those conversations, muddies the waters, and slows scientific progress.

Practical Suggestions

Ask yourself whether/when the terms "structural" or "reduced form" are needed.

- usually one can just say what the assumed model is, what the objects/quantities of interest are, how these are identified, and how one estimates them.
- when you only want to convey features or suggestive patterns of the data, label this as descriptive.
- when you want to estimate a simple model as a first pass (recognizing that this model requires assumptions you might not want to defend), just say this.

When others speak/write less precisely, pause to think carefully about what they really mean.

The Role of Economic Models in Empirical Work

A Case for Empirical Models Derived from Economic Theory

Going beyond data description requires a model.

The model need not reflect any economic theory, but I will argue that often it should—that modeling the behavior generating the observables often provides valuable guidance for empirical work.

Some Roles of Economic Models in Empirical Work

As economists we often

- have questions that can be properly posed only within an economic model, e.g.,
  - what is the short-run elasticity of demand for Uber rides?
  - how much higher are health care costs in the US as a result of moral hazard? adverse selection? hospital mergers?
  - what share of the male-female wage gap reflects discrimination?
  - what are the equilibrium distributional implications of liquor taxes?
  - what effects would a $XX carbon tax have on carbon emmissions and GDP?
- are willing to explore what can be learned by maintaining certain assumptions on the underlying preferences, technologies, equilibrium concepts, etc.

Economic Models in Empirical Work

Many important questions can be answered only by exploiting economic models to provide a logical framework for interpreting the data:

- tell us what to look at: what are the structural features of interest for our questions? A strangely controversial opinion: one should not define the object of interest based only on what some statistical procedure produces!
- define what it means to have a "valid" estimation method—one that uncovers the features of interest rather than something else
- provide functional/causal/probabilistic relationships that can be used to estimate of the structural features of interest; e.g.,
  - optimality/best response conditions that relate observables to primitives
  - IV conditions.

The Role of Economic Models, A Cautionary Tale: Part 1

A researcher observes price P and quantity Q of a good in many markets

- He says: "I do not want to impose arbitrary restrictions from a model. I just want to let the data speak."
- He regresses Q on P, finds a positive correlation, and concludes "Initial evidence suggests that P has a positive effect on Q."
- He then adds some covariates Z to the regression and obtains similar results. The researcher concludes, "The positive effect of P on Q is robust to the inclusion of a rich set of controls."

(So far, nonsense.)

A Cautionary Tale: Part 2

- The researcher now considers the use of instrumental variables methods, characterizing this either as a robustness check on the original analysis, or as a way of controlling for possible "confounds"
- The researcher suggests that a measure of the availability of a substitute goods be used as an instrument for P. TSLS now reveals a stronger positive "causal effect" of P on Q. The researcher concludes that the original results are qualitatively robust, but that controlling for endogeneity of P eliminates a downward bias in the OLS estimates.

(still nonsense)

A Cautionary Tale: Part 3

- Another researcher reads the paper and has a different idea for an instrument: the manufacturing wage in the local market, something also left out of Z that plausibly affects prices.
- TSLS now yields a precisely estimated *negative* “causal effect” of price on quantity
- The researcher concludes, “the causal effect of P on Q is heterogeneous. The effect one measures depends on the set of price changes created by variation in the instrument one uses.”

(more nonsense).

How Do We Know What Is Nonsense?

Our researcher started with a common approach: he (correctly!) conjectures that a variable X may affect another, Y, and explores the relationship with regression, with or without IV, interpreting at least the latter as "the causal effect of X on Y."

Where did he go wrong? How do we know that something has gone wrong?

A Nose for Nonsense?

Here we know something is wrong because we already have a deeply ingrained idea of how prices and quantities are determined in a market. We automatically have at least elements of an economic model in mind.

Here intuition from elementary economics is enough to tell us that the very notion of "the causal effect of P on Q" is not well defined!

In general, however, your intuition may fail you (even for supply and demand: see, e.g., Berry-Haile 2021). Failing to write down an economic model that justifies your empirical approach will often lead to an error. Examples (some shockingly close to the cautionary tale above) in the literature abound, even from famous economists.

Economic Models: Tools for Avoiding Nonsense

Once the researcher suspects that there may be omitted factors, selection, endogeneity, etc., *some type of model is necessary* to determine what "fixes" will work (indeed, what "work" even means). In the example a model is needed...

- to recognize demand and supply as distinct objects
- to define what objects should be measured
- to recognize which structural errors are present (e.g., must be held fixed to measure a demand elasticity)
- to define what it would mean to have a valid estimation approach or valid IV (these often appear when one writes down a model).

The Role of Models

Stepping Back

"all models are wrong, but some are useful." – George Box

Useful at a minimum because without a model of some kind there is typically only hand waving. Attempts to go beyond data description without a model are "not even wrong" — i.e., one cannot even define what "right" means.

The Role of Models

Stepping Back

"Art is not truth. Art is a lie that makes us realize truth..." – Pablo Picasso

The art of empirical work includes selecting a useful model *for the purpose at hand*. This will allow one to justify an interpretation of a measurement—a minimal requirement for a "correct" answer. The model will always involve assumptions that one could question, debate, reject, or improve upon. Others may view a different model as more useful.

But only by specifying a model can one speak coherently about whether the maintained assumptions are problematic, whether certain data allow measurement of the quantities of interest, what alternative assumptions might imply, and how science might progress.

Summary

1. Don't be fooled by common (ubiquitous!) abuses and confusion about key terminology in empirical economics. Try not to become part of the problem.
2. Better yet, look for opportunities to overcome the false barriers and bring together insights from artificially disconnected literatures.
3. Don't underestimate the extent to which an economic model can be useful—even essential—to good empirical work.

Fun Reading

Koopmans, T. (1947). "Measurement Without Theory," Cowles Foundation Discussion paper 25a. http://www.jstor.org/stable/1928627

Koopmans, T. (ed) (1950). *Cowles Monograph 10. Statistical Inference in Dynamic Economic Models*, London: Wiley. (esp. chapters by Hurwicz, Koopmans) http://cowles.yale.edu/sites/default/files/files/pub/mon/m10-all.pdf

Marschak, J. and W.C. Andrews Jr. (1944), "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143-205., http://www.jstor.org/stable/1905432