

Milton Friedman — and Others — on Identification*

(Started: December 15, 2013; Revised: July 21, 2014)

Some thoughts about identification and empirical work in economics. This started with a comment Tom Sargent made at a seminar, that Milton Friedman had some interesting things to say about identification. I tracked them down in “The methodology of positive economics,” the first chapter of *Essays in Positive Economics*. Section 1 is my summary. After I wrote that, Allan Collard-Wexler sent me a link to a talk by David Card and Tom sent a draft of a book review by John Rust, so I worked them in, too. As a result, this kind of meanders, but I think the content, if not the presentation, has some inherent interest.

No guarantees that any of this makes sense — it’s simply my attempt to think through the logic of empirical work in economics. Comments welcome on any or all of this.

Warning: I’ve streamlined the quotations to make them read better.

1 Friedman on positive economics and identification

This is my summary of Milton Friedman’s comments, but let me know if you think I missed something.

On positive science (page 7):

The ultimate goal of a positive science is a theory or hypothesis that yields meaningful predictions. Such a theory is, in general, a mixture of two elements: a language or filing system to promote organized thought and factual evidence that shows whether this language is useful for analyzing concrete problems.

On supply and demand (pp 7-8):

The example of supply and demand illustrates the approach. Viewed as elements of the language of economics, they provide categories of factors affecting prices. Their usefulness depends on the empirical generalization that the forces affecting demand and supply will yield two lists that contain few items in common.

This generalization is valid for markets like the final market for a consumer good. In such a market there is a clear distinction between the economic units that demand the product and those that supply it. There is seldom much doubt whether a particular factor should be classified as affecting supply or demand. But the generalization is not always valid. For example, it is not valid for the day-to-day fluctuations of prices in a speculative market. Here almost every factor can be classified with as much justification under the heading “supply” as under the heading “demand.” These concepts can still be used, but they are less useful because they have no meaningful empirical counterpart.

*Working notes, no guarantee of accuracy or sense.

Comment: It's not hard to interpret this last example in the context of a rational expectations model, where supply and demand get scrambled.

On the multiplicity of hypotheses consistent with evidence (pp 11-12):

The validity of a hypothesis in this sense is not by itself sufficient. Observed facts are necessarily finite in number; possible hypotheses, infinite. If there is one hypothesis that is consistent with the available evidence, there are always an infinite number that are. ... Additional evidence may rule out some of these possibilities; it can never reduce them to a single possibility alone capable of being consistent with the finite evidence. The choice must to some extent be arbitrary.

Comment: This sounds like identification, but I think he's alluding to something deeper. We continue...

On the Cowles Commission's work (pp 12-13, incl fn 11):

Empirical evidence is vital at two different stages: in constructing hypotheses and in testing their validity. Full and comprehensive evidence is needed to assure that a hypothesis explains what it sets out to explain — that its implications are not contradicted in advance. Further testing involves deducing new facts and checking them against additional evidence.

Economists associated with the Cowles Commission have divided the step of selecting a hypothesis consistent with known evidence into two substeps: first, the selection of a class of admissible hypotheses (the choice of a “model”); second, the selection of one hypothesis from this class (the “structure”). This subdivision may be heuristically valuable [but] from a methodological point of view it is entirely arbitrary.

One consequence is the so-called “identification” problem. ... It may be that the evidence used to select the final hypothesis from the subclass can be consistent with at most one hypothesis in it, in which case the “model” is said to be “identified.” But this way of describing the concept of “identification” is essentially a special case of the more general problem of selecting among alternative hypotheses consistent with the evidence — a problem that must be decided by some such arbitrary principle as Occam's razor.

My take: He's saying, I think, that the Cowles Commission trivialized the process of selecting one of many evidence-consistent hypotheses by restricting the choice set. In applications, there are always alternatives that go beyond this. He concludes fn 11 this way: “However useful the two substeps may be in some contexts, they raise the danger that different criteria will unwittingly be used at the two stages.” In short: They solved a narrow technical problem (beautifully, I would say) but didn't really address all the possible alternative models that might be consistent with the evidence.

2 Rust, Sims, and Card on identification

Colleagues sent links to related comments by David Card and John Rust. Both do work well outside my own areas of research, so I don't guarantee I get all of this, but here's my take. I've added some comments by Chris Sims, at his best as a critic.

Single-equation methods and instrumental variables. Rust and Sims raise the issue of single-equation methods, particularly instrumental variables. Rust's lovely rant includes this comment:

Levitt exemplifies recent generations of MIT graduates who have been taught that the only tools one really needs in order to be a successful applied economist is a little knowledge of regression and instrumental variables.

See [here](#) or [here](#). Rust argues, I think, that you need more than a “magic instrument” to make an interpretation of the evidence persuasive.

Sims has a delightful rant of his own in his Rudebusch comment, which includes

The identified VAR literature has begun attracting vigorous criticism from economists comfortable with the currently fashionable view that macroeconomics, properly executed, never requires thinking about more than one regression equation at a time.

See [here](#). His comments on Angrist and Pischke expand on the same idea:

Recent enthusiasm for single-equation, linear, instrumental variables approaches in applied microeconomics has led many in these fields to avoid undertaking research that would require them to think formally and carefully about the central issues of nonexperimental inference. ...

The Ehrlich work on capital punishment [does capital punishment affect murder rates?] is a good example. I read that work with interest at the time. ... What made me most uncomfortable was that it assumed a list of exogenous variables without discussing why they were both plausibly exogenous and, probably more important in that case, why the pattern of exclusion restrictions on those variables was reasonable. ... But using instrumental variables while simply listing the instruments, with little or no discussion of what kind of larger multivariate system would justify isolating the single equation or small system to which the formulas are applied, was, and to some extent still is, a common practice.

See [here](#), pp 59-62.

Both comments fit, I think, in the Cowles Commission tradition: identification in simultaneous equations models involves the other equations. So what are they? Without that step of thinking “formally and carefully” about the rest of the model, it's not clear what we end up with.

Identification and alternative theories. Sims and Rust go on to echo Friedman on the broad range of alternative theories that might account for the evidence. Sims (pp 60-61) again:

The same data generally are subject to multiple interpretations. It is not that we learn nothing from data, but that we have at best the ability to narrow the range of substantive disagreement.

Rust's review of Ken Wolpin's book ("Limits of inference") includes these comments (page numbers in parentheses):

There are a number of inherent limits to inference that may be insuperable. (4)

I think structural econometricians need to think more deeply about whether they can justify whether *any* parameters are really "structural" in the sense of being policy-invariant. (9)

The curse of dimensionality forces us to work with fairly simple models because we can't solve bigger, more realistic ones. (10)

The most daunting limit to knowledge is the *identification problem* of trying to infer the *structure* — [the] set of *primitives* that imply a probability distribution for the observable variables. Structural models depend on *maintained assumptions* that are outside the domain of the identification analysis. ... [For example], most structural models allow for unobserved state variables. Common maintained assumptions are *additive separability* (the unobserved state variables enter the utility function in an additive way) and *conditional independence* (the unobserved state variables ... do not directly affect the distribution of observed state variables). [Even this may not be enough.] (10-11)

I believe that most interesting economic models are either non-parametrically unidentified or at best partially identified. (14)

Like Sims, he's suggesting there's no unique path to identifying the primitives underlying the distribution of observables. Note, too, that he goes beyond the canonical simultaneous equations setup to include the possibility that there are some things we don't observe. That's pretty much always the case in applied work. The macro version of this was laid out by Hansen and Sargent (JEDC, 1980) in a relatively simple one-dimensional example of a dynamic optimization problem. Similarly, in lots of modern work in IO, theoretical structure is used to overcome lack of data about some of the key variables.

The role of structure. Rust starts with Koopmans' famous [measurement without theory](#) critique atheoretical business cycle research. Here's a longer version of a Koopmans quotation (p 172) he uses:

The [Burns and Mitchell] book is unbendingly empiricist in outlook. Granted this basic attitude, it shows great perseverance and circumspection on the part of the authors in handling a vast amount of statistical data. ... But the decision not to use theories of man's economic behavior, even hypothetically, limits the value to economic science and to the maker of policies.

I'd say Burns and Mitchell look pretty good in retrospect. They established some properties of business cycles, and any theory of their time is unlikely to be as useful to us as this has been. Koopmans (p 161) makes a similar point in passing:

Burns and Mitchell are more consistently empiricist than Kepler was. The latter made no secret of his predilection for the principle of circular motion until observations spoke decisively for the elliptical orbit. He held other speculative views as to the role of the five regular solids and of musical intervals in the proportions of the planetary system.

He's arguing that Kepler was more theoretical than Burns and Mitchell, hinting this is a good thing, but his examples of Kepler's theories suggest instead they weren't all that helpful. Although maybe an incorrect theory (circular orbits) was helpful in thinking about the evidence.

But let's consider the role of structure in empirical work. Surely structure is good if it's right, but how do we know that? Sims is agnostic:

[Econometricians] should be able to fit loosely interpreted models that characterize patterns in the data, to impose identifying restrictions that allow richer interpretations, to compare the fit of alternative sets of restrictions, and to describe uncertainty about results, both in parameter estimates and across models.

Hard to argue with that. But Rust is pessimistic about ultimate success:

[I]t is fair to ask whether structural models really have succeeded, and resulted in significantly more accurate and reliable policy forecasting and evaluation. I think the jury is still out on this, because even though Wolpin has offered some compelling examples of successful use of structural models for policy evaluation, there are still relatively few clear-cut successes where structural models have had a clearly measurable positive practical impact on policymaking.

It seems we're a long way from Koopmans' vision of theory guiding empirical work. That said, John has been a strong advocate for structural econometric work. My own take (next section) is that models are helpful in giving us precise interpretations of the data, but we always run the risk that we don't have the best interpretation. Science works that way: we come up with better interpretations all the time.

Card's perspective. Card, of course, is an expert in the kind of thing Rust and Sims are complaining about. My knowledge of his work is hearsay at best, but I've watched a [video](#) of his Woytinsky lecture and scanned the [slides](#). I recommend the video, it's terrific.

Some of his points:

- Causality. Two common views of causality: model-based (theoretical model tells us what's exogenous and what's endogenous and how they're connected) and design-based (manipulations of x lead to changes in y). Neither is enough on its own. Astronomy has no design elements (what's the manipulation?) and medicine often has little in the way of a model (how does aspirin work?). [Haavelmo](#) (Econometrica, 1944) says we need both.
- Models. All papers in economics have models, but they vary in how explicit they are. McFadden's work on discrete choice is explicit. Levitt's work on police and crime is implicit. Ditto much of Card's work. One of the challenges of work with implicit models

is extrapolation: you estimate the impact of a given manipulation, but can't say much about others.

- Identification. Econometric identification: exclusion restrictions, classic Cowles and modern variants. Substantive identification: ruling out alternatives, more art than science.
- Welfare. "I've always been uncomfortable with this, but you need a model to do it."

My summary: We often don't know much about the model, but we can learn things in a relatively robust way from natural experiments.

3 Approaches to empirical work

What does this have to do with me? I'm not sure. I'm a macroeconomist, and in my field I see several useful steps in the empirical research program that constitutes much of mainstream macro. Good work can take any of these forms:

- *Facts*. Summaries of data properties — what you might think of as exploratory data analysis. The idea is to establish some basic facts a line of research is attempting to explain. Card (1:17) echoes this at the end of his talk: "There's a need for many different kinds of applied work. ... There's a lot of room for what I would call descriptive work, where you pull the facts together."

Obvious examples: means and standard deviations, correlations, cross-correlation functions, principle components, and so on. In macro, we might list dynamic factor models (Sargent-Sims, Stock-Watson), (bivariate) cross-correlation functions (Hodrick-Prescott), and Sims's work with VARs. Burns and Mitchell's work on business cycles is similar in spirit, although we now have better tools.

One interesting feature of much of this work is that there's no distinction between endogenous and exogenous variables, we treat them all the same. Sims's [discussion](#) of Angrist and Pischke has some of this flavor. He notes, with regard to Ehrlich's work on capital punishment, that there's a lot of variation in murder rates across states, little of it connected to capital punishment. That suggests we might think about where this variation comes from before we get bogged down arguing whether capital punishment deters murder.

- *Screening tests*. Establish essential features of the data and use them as (in Richard Rogerson's terminology) "screening tests." If a model doesn't have these features, we move on to the next one. There's not a clear separation between this bullet and the previous one, but I had in mind here a modest use of theory that focusses our attention on specific features of the data.

Examples: (i) Rogerson has in mind volatility and comovement in business cycle models: investment is more volatile than output, and consumption less volatile, and all of them are strongly and positively correlated. The theory content here is our understanding of consumption smoothing from earlier work, which is built into most macro models. (ii) Another example is measurement based on an aggregate production function. Whether we think of it as a sensible object or not, it's a useful measurement tool for things like growth accounting and the Solow residual. We can ask of models whether they generate residuals like those we see in data. (iii) Shimer's work on search models. (iv) Perhaps my favorite

example is the Hansen-Jagannathan bound. Lars Hansen has a nice description of the general approach: “You can do something without doing everything. Suppose you’re interested in the connections between financial markets and the macro economy. In order to do that using statistical methods, you have to have a formal model, and to do that requires a whole lot of inputs. So the question is how you do that without having a full-blown model of everything.” See [here](#).

In macro we’re often stuck trying to find models that pass basic screening tests. That is, we have models with some obvious flaws, but hopefully provide some insight nevertheless. This is a step prior to Friedman’s suggestion that we have lots of models that account for the evidence, but it’s where we are right now. To paraphrase Rust, we’re not looking for truth, we’re looking for approximations that capture some of the obvious features of the data. That’s turned out to be hard enough.

Question: Can we think of design-based methods like this? Would Card argue that he’s reporting some features of the data without being explicit about the rest of the model?

- *Models*. I think of a model as an interpretation of the data. BLP strikes me as in this spirit: a concrete interpretation of price and quantity data for cars.

This kind of work can be very helpful in overcoming lack of data and in extrapolating to policies well outside what we’ve observed. Antitrust applications in IO, for example. But as Rust says, anything of this sort is founded on maintained assumptions that deserve a closer look. Card puts it this way (1:00:45): “They’re showing that a model can describe reality and then using it to illustrate what would happen in an alternative reality that doesn’t exist.”

- *Identification*. Can we tell alternative models (interpretations) apart? How? Card calls this *substantive identification*: “ruling out alternative explanations for the data (getting past the identification police).” In applications, we often worry about alternatives based on reverse causality, selection, or unobservables. As Rust says, it’s not clear we’re capable of this. Or as Sims says: there are typically multiple interpretations of the evidence. Any model, then, is simply one possible interpretation.

But I’d say that’s what we do: we give interpretations to what we see. We’ll never know if we have the only or the best interpretation, but it’s part of the process of understanding economic behavior. I think that’s true outside of academics, too. Here’s a line from Susan Athey about data scientists in the business world:

I think when you have large amounts of data, ... you have a greater ability to let the data speak, and so you can be much less reliant on assumptions. But you still need a strong conceptual framework to understand what’s coming out. Among data scientists, the ones who can define a question and introduce a new way of looking at the data — those data scientists are rock stars. They’re pursued by every company and they move up the hierarchy very quickly. And there are never enough of them.

See [here](#).

All of which reminds me of a comment Lucas made years ago about the assumption of policy-invariance: some things are good bets, some are not. Rust and Card would agree,

I think, that it's not clear there are any invariant deep structural parameters, but I still think there are some good bets out there.

So where does that leave us? To quote Rust (p 17) again:

[M]odels are necessarily highly oversimplified approximations to reality and can never be correct. "Econometrics as a search for truth" may be too idealistic a goal, given the limits to inference that we face. It might be better cast as a "search for models that provide reasonably good approximations" to complex phenomena.

Sounds right to me. I'd add that it would be helpful if at least some empirical work could be done in a way that's somewhat model-independent, so that features of the data can be established that would be useful even if someone in the future has a different theoretical model in mind. Given the uncertainty about what kind of model is appropriate, there's clearly value in separating the search for facts from the search for interpretations of them.