

What would a scientific economics look like?

Peter Dorman [Evergreen State College, USA]

Copyright: Peter Dorman, 2008

Abstract

Sciences are loosely characterized by an agenda to describe the mechanisms by which observable outcomes are brought about and the privileging of propositions that have been demonstrated to have negligible risk of Type I error. Economics, despite its pretensions, does neither of these and should not be regarded as scientific in its current form. Its subject matter, however, is no more recalcitrant to scientific procedures than that of many other fields, like geology and biology. The benefit of bringing economics into greater conformity with other sciences in its content and method would be twofold: we would be spared the embarrassment of unfounded dogma, and over time economics could assemble an ever larger body of knowledge capable of being accepted at a high level of confidence. A scientific economics would take Type I error far more seriously, would study mechanisms rather than a succession of states, would be more experimental and would attach greater value to primary data collection.

This is a trick question, of course. Loaded into it are many assumptions: that we know what “science” is, that the economics we see today is not scientific in this way, that it could be, and that it would be beneficial if it were (or at least interesting to speculate on). After you think you know all of this, the title question does not add so much.

But if it has captured your attention, it has done its job. So now I will briefly flesh out these assumptions and conclude with a vision of the type of economic practice that comports with my idea of science.

What is science?

This is not the occasion for a detailed exposition of a philosophy of science that situates itself within the enormous literature that has grown up around this topic. Let’s take a step back and see what common features characterize most scientific work and distinguish it from other cultural and intellectual endeavors. I see two, an agenda based on providing at least potentially measurable causal mechanisms to explain why we observe the world we do, and a privileging of claims that have been shown to bear a negligible risk of Type I error (see below). In addition, of course, scientists do lots of things the rest of us do: compete for resources, bolster their guilds and professional perquisites, engage in the persuasion of their peers, and so on. What is interesting, however, are the two characteristics that make scientific pursuits distinctive.

Minimization of Type I error

Recall that this is the mistake of affirming a proposition when it is false, while Type II error is failing to affirm it when it is true. Nearly everything we associate with scientific protocols is intended to identify factors that might confound empirical tests and lead to Type I error. This includes the failure to fully document all the properties of the experimental apparatus, protocols of experimental method, setting a low threshold for acceptable p-values and the role of transparency and replication. From a practical point of view, in “real” sciences, to announce a result and then have it shown to be in error due to some mistake in procedure is to risk the end of one’s career: minimizing Type I error is taken very seriously. By comparison, failing to recognize that the data support a different hypothesis is a sign of dullness or distraction, but the researcher who errs in this respect will experience it as a setback, not a career-breaker.

There is a cultural aspect to this obsession with one sort of error at the expense of the other. At some time in the past various thinkers and students of nature developed a powerful desire to separate those things we truly know from the many others we only suspect. This can be seen in the geometric proofs of the ancient Greeks and the development of formal experimental methods much later. Of course, at any point in time many propositions are in limbo, having gained some credence but not (yet) passing into the circle of those for which the risk of false acceptance is negligible. Scientists spend a lot of time debating these propositions, but when they do they are not acting very differently from people in other walks of life who also engage in disputes. What is special about science is that it is expected that these debates will eventually end when the risk of a false positive is either dispelled or confirmed.

But there is also a purely practical aspect to the practice of attaching a very high cost to Type I error and a much lower cost to Type II. Science has evolved into a vast enterprise based on an elaborate division of labor. Experimental protocols require that most potential confounders, of equipment, sampling, or other aspects of research design, be identified and their effects measured with near certainty. This entails dependence on the results of other researchers. If one widely accepted claim later proves to be false, it may call into question a whole chain of subsequent studies. Type II error slows the progress of science; Type I error forces science to backtrack to some earlier state.

The search for causal mechanisms

The question of whether antecedent condition x can be said to cause subsequent outcome y has proved vexing for philosophers, but I am interested in a much simpler matter. In nearly every science, attention is given primarily to the process by which changes take place, or more precisely the mechanisms that can be seen to operate during such a process. This is true even if this knowledge is not sufficient for much predictive success. What does a biologist, for example, study when she studies a fish? It could be the mechanisms by which different organs function, or it could be the life cycle attributes that explain why certain environmental factors influence population size, or it could be the genetic pathways that connect current taxa to their evolutionary antecedents. In any case, the result is an accumulation of factual knowledge about fish in their various contexts of time and place, resting on the mechanisms by which they function and evolve. There is no “general theory of fish”, although there are general properties that biological mechanisms have to obey, given by thermodynamics, the chemical mechanisms (drawn from another science) that underlie cell growth, and the constraints of nutrient availability and transport.

Economists have been seduced by a different vision of science, with its roots in those same ancient geometric proofs: that the foundation of science rests on a deductive theory, where “explanation” means “producing a story that can itself be expressed as a deduction from top-level theory”. This is a dream of mathematics and much of physics. It is not characteristic of scientific work in general, however, and its usefulness is limited in fields of research that are extremely complex and dependent on a vast array of contingent factors. Geology would be such a case, where certain general concepts (like plate tectonics) are broadly accepted, but what matters in practice is the knowledge of concrete mechanisms, such as what forces are at work in the subduction of plates or, on a more mundane level, the movement of subsurface water through various soil and rock strata. Geologists cannot give you a general theory of earthquakes, but they can describe rather accurately the process by

which the force generated by plate collisions is stored and transmitted in specific formations, and the same can be said for the skills of the hydrologist you might consult before deciding whether you ought to build your house on a particular slope.

Causal mechanisms are the preferred subject matter of science for several reasons. They appease our curiosity. They are more likely to be testable in ways that minimize Type I error than process-independent claims about prior and subsequent states. Above all, they provide knowledge that is frequently useful even when it is incomplete—as knowledge invariably is in complex domains. If you know some mechanisms but not all of them, you still know something about how a system works.

Lip Service

Economics gives lip service to both of these distinguishing characteristics of science, but little more. In practice, the adjective “scientific” is given to work that rigorously adheres to deduction, not the rooting out of Type I error or the identification of mechanisms.

The false promise of econometrics

Don't get me wrong: I love econometrics. I practice it, enjoy it and learn from it. As it is presently constituted, however, it makes only the weakest attempt to avoid Type I error. Two weak gestures can be observed. First, practitioners reject all results whose p-value is not sufficiently low. Second, if they are conscientious, they search for estimation methods that are suited to the data they are analyzing. This shows up in discussions of whether multinomial regressions should be ordered or not, or what identification strategies are likely to be available, or whether fixed effects can be introduced. This is all well and good, but it is not the same as minimizing Type I error. Such a goal would require identifying every possible confounder, whether they take the form of missing variables, assumptions on functional form, or the use of theoretical priors that, while conventionally accepted, have not themselves survived testing designed to minimize Type I error. You will be hard-pressed to find a single econometric study that meets this criterion, but you will find many studies in “real” sciences that do. (The biggest problem in those sciences is not the elimination of Type I error in the individual study, but the extent of external validity which may have been sacrificed to achieve purity in research protocol.) What you will typically find are studies that essentially calibrate models whose general structure and content are not put into question. What counts as a test in such research is the ability of the model to achieve calibration: if you can do this, the model is said to “be consistent with” the evidence. Of course, the number of models capable of being empirically calibrated is much larger than the number that would survive rigorous testing based on minimization of Type I error.

What makes this failure so pernicious for economics is that the entire edifice is built on prior results that are themselves at great risk of being false positives. And it is interesting that no one much cares. This point is fundamental.

Half an equilibrium

Perhaps the best way to tell this story is in the order I became aware of it. It struck me that prominent economists were discussing identity relations, such as those that form the basis of macroeconomic accounting, as if they were functional. What motives, they asked, were

causing some agents to lend enough to meet the (given) borrowing demands of others? The explanations they gave left open the possibility that, over protracted z periods of time, the financial identities could be violated. I discussed this elementary error in some detail in an article I wrote a year ago on global imbalances. (Dorman, 2007)

At first blush, this seemed to be a careless mathematical mistake, dropping the third bar from the identity sign, perhaps due to writer's cramp. (Even in the age of computers, much math is done longhand.) Indeed, if you look through many a journal article you will see only equations, no identities, even if the subject is macroeconomics and identities are very much in play. It may be that the habit is picked up in graduate school, and that many otherwise well-versed economists have never encountered the idea that three bars are not the same as two.

Eventually, it dawned on me that the difference between identity and functional relations vanishes if one considers only equilibrium states, in which functional relations, by definition, hold. And this is, in fact, the methodology of nearly all modern professional economics. One specifies an equilibrium, tweaks the parameters, and predicts what the new equilibrium will be. Or, if the exercise is econometric, it may simply be a matter of "testing" the model by estimating the parameters that would generate the observed pattern of outcomes in equilibrium. In any case, one is either in an equilibrium or undergoing change from one to another the way teleportation works in science fiction stories; in either case the distinction between identity and functional (or behavioral) relations can safely be ignored.

Yet, as all of us learned somewhere along the way, a stable equilibrium requires an out-of-equilibrium process that draws us in. It is interesting to note that stability becomes a necessary consideration when more than one equilibrium can arise from a model; typically there are unstable equilibria that separate their stable cousins. But most economists shun such models (although they may well be more descriptive of real-world phenomena, as I argued in Dorman, 1997), and in any case, the prevalent methodology considers primarily the equilibrium state and seldom the equilibrating process. Since identity relations constrain behavioral adjustments in the course of equilibration, but are indistinguishable at the equilibrium itself, their special character drops from sight.

Thus the ultimate cause of error is not random carelessness, but the limited attention economists give to mechanisms rather than end states. The irony, as we shall see shortly, is that mechanisms are usually much easier to observe and measure with confidence, and may well give us the sort of information that more usefully informs practical decision-making.

Scientific economics is possible.

Whenever the criticisms from philosophers and other professional methodologists become too severe, we hear the excuse that economists just can't do the sort of things other scientists do. We can't do experiments the way they can. Our subject matter is more complex. People are unpredictable in fundamental ways.

These appeals are unconvincing. Geologists can't do experiments on many of the questions that concern them, nor can ecologists, nor evolutionary biologists. Ecology is horribly complex, and so, we are learning, is climate science. Many units of observation and analysis, ranging from micro-organisms to tectonic plates, behave in ways that are

unpredictable on the basis of present knowledge. Science is very difficult, and what is not known dwarfs what is.

But economists give themselves too hard a job. It is indeed extremely difficult to characterize equilibrium states in sufficient detail to generate true Type I error-minimizing tests, while it is much easier to identify and test for mechanisms.

Perhaps the simplest and most universal example will make this clear. Consider the standard supply-and-demand diagram. The professor draws this on the chalkboard, identifies the equilibrium point, and asks for questions. One student asks, are there really supply and demand curves? Where would you go to look for them? Ah, it's not so easy, comes the reply. Yes, in principle these curves exist, but they are not directly observed in nature. You can do market research in which you ask a sample of consumers how much they would buy at various prices, and this could give you an estimate of the demand curve, but of course there would be a certain amount of error in the process. And the supply curve is even more difficult. We will see in another week that this is derived from the marginal cost schedule, but in practice firms often find this difficult to calculate with accuracy. And even worse, in another week after this we will find out that, if competition is not perfect and firms behave strategically in the market, there is no supply curve at all.

If the answer stops here, the students will be left wondering why they are studying such a useless theory. But we know there is another way the answer might proceed. The professor could say, the supply and demand curves are only for the purpose of organizing our thoughts; they are not "real" in the way you are asking for. But we can use them to identify two other things that are real, excess supply and excess demand. We can measure them directly in the form of unsold goods or consumers who are frustrated in their attempts to make a purchase. And not only can we measure these things, we can observe the actions that buyers and sellers take under conditions of surplus or shortage.

In this easiest of cases, it is already clear that mechanisms are more susceptible to empirical methods than models of endpoint (equilibrium) states. This observation applies with greater force as we move toward ever more-complex forms of equilibrium modelling. Fortunately, the antidote is beginning to emerge in such areas as labor market search theory and behavioral finance, which have brought concrete mechanisms back into the picture. As these fields develop, the more general models of their infancy give way to diverse findings across particular market segments, cultures and contexts. And that's what we should expect: there is no general theory of fish either.

Scientific economics would be better than what we now have.

This is actually the most difficult case to make. In some ways the point is obvious. For instance, economists bend their research toward axiomatic theories that are almost embarrassing in their pre-scientific naiveté. Consider utility theory, for instance, which is now taking a drubbing at the hands of experimental psychology and neurophysiology. A scientific orientation would free us of such vestigial dogmas.

The more difficult issue concerns the relationship between science and policy. Economics is never more than a few centimetres away from significant matters of human well-being, and the criteria for policy are strikingly different than they are for science. For

questions of policy both types of error are potentially costly, and those who offer advice must balance the risks of false positives and false negatives based on the relative consequences of each. This is how I interpret the parable given by McCloskey (2002), in which you are out in the world and, amid a confusion of noises, think you might be hearing someone crying out "Help! Help!" But you are not sure, it could be a discussion about seaweed and someone is making the point "Kelp! Kelp!" In this case you consider that the cost of not running over to be of assistance if it is needed far exceeds the cost of running over if it is not: the cost of Type II error trumps the cost of Type I. This is the logic of policy but not science.

The distinction between scientific and policy perspectives on error is exploited by those who benefit from inaction. Why act on climate change or similar threats if scientists cannot exclude the possibility that the whole matter stems from false positives? This confuses two different sets of criteria, and policy-aware scientists know that the standards for certainty in one domain are necessarily different from the standards for action in another. Thus the Intergovernmental Panel on Climate Change places percentage confidence estimates on its various predictions; 80%, for instance, earns a "high" even though a p-value of .20 would flunk every known test of statistical significance. (IPCC, 2007)

But if the main purpose of economics is to guide human actions in economic affairs, and if the criteria for this guidance differ from those of science, why should economists try to be more scientific? This is a fair question, but it should be remembered that practical considerations have always been important in the most scientifically respectable disciplines as well. In fact, one could say the field of technology, broadly understood, constitutes an entire universe, side-by-side with science, in which Type II error matters quite a lot. This has not escaped the notice of those who fund scientific research, and there are frequent spats over how valuable is the "ivory tower" work in which Type II error is given little if any weight.

This comparative perspective suggests that a scientific economics could justify itself along the same lines that other sciences have in the past. In my view, two arguments are strongest. First, there is the familiar appeal to serendipity: sometimes you have to separate yourself from practical concerns in order to free the imagination to develop new practical applications. The ruthless pursuit of what can be known with near certainty forces the scientist to take seriously many possibilities the technologist might overlook. (And the opposite is often true as well, of course.) Perhaps most will prove to be dead ends, but a few may open the doors to entirely new ways of thinking about problems and their solutions.

The second, and much the more powerful, concerns the long run. Over time, a scientific enterprise that minimizes Type I error will accumulate a body of knowledge and methods on which ever more productive research can take place. This foundation will be available equally to the policy researcher, who will then be able to generate more powerful tests that reduce the trade-off between the two types of error. This long term symbiosis can be seen, for instance, in the fruitful relationship between academic toxicology and epidemiology, which cautiously shun the risk of false positives, and hazard assessment as conducted by regulatory agencies, whose mandate places far greater emphasis on the risk of false negatives. Today's hazard assessment is more reliable because of generations of accumulated advances by researchers whose scientific criteria would not have been optimal for the assessment of exposure risks at any single moment in time.

One could contrast this with many branches of economic research that have hardly advanced at all, due to the lack of interest in the potential for error. Conspicuous in this

respect is CGE (computable general equilibrium) modelling, which is never subjected to serious retrospective testing. There exists no evidence I am aware of that establishes the extent, if any, to which such models have improved our ability to make forecasts. The models become more elaborate and some of their components are calibrated more precisely, but there is no reason to believe that their effectiveness as analytical tools is greater now than 20 years ago. This illustrates by its absence the role that the systematic effort to minimize Type I error plays in establishing the progressive character of science. (CGE modelling also exemplifies the failure to consider mechanism, which is why the devastating Debreu-Mantel-Sonnenschein results have been completely ignored; for an intuitive explanation, see Dorman, 2001.)

So what would a scientific economics look like?

I have mostly answered this already: it would look like other sciences whose objects of study are complex, heterogeneous and context-dependent. It would study mechanisms primarily and end states only for heuristic purposes. It would be predominantly empirical, where this encompasses both statistical work and direct observations on economic behavior (which may also entail statistical analysis). It would ruthlessly identify potential sources of Type I error and strive to eliminate them in hypothesis testing. Experimentation, in the lab and in the field, would become more common, but even more important, primary data collection of all sorts would be accorded a very high value, as is the case in all true sciences. Its macro models would come to look like macro models in hydrology or biogeochemistry: simultaneous differential equations representing mechanisms rather than static end states embodying (a single) equilibrium. Economists would increasingly find it useful to collaborate with researchers from other fields, as their methodological eccentricities are abandoned. Finally, there would be a much clearer distinction between the criteria governing scientific and policy work, insulating the former from some of the influence exerted by powerful economic interests and freeing the latter to adopt an ecumenical and risk-taking approach to tackling the world's problems.

References

- Dorman, Peter. 2007. "Low Savings or a High Trade Deficit: Which Tail Is Wagging Which?", *Challenge*. 50(4): 49-64.
- Dorman, Peter. 2001. "Waiting for an Echo: The Revolution in General Equilibrium Theory and the Paralysis in Introductory Economics". *Review of Radical Political Economics*. 33(3): 325-33.
- Dorman, Peter. 1997. "Nonconvexity and Interaction in Economic Models". Social Science Research Network Working Paper No. 1232. *Microeconomic Theory* 2(2).
- IPCC. 2007. *Climate Change 2007: Synthesis Report*.
- McCloskey, Deirdre N. 2002. *The Secret Sins of Economics*. Chicago: Prickly Paradigm Press.

SUGGESTED CITATION:

Peter Dorman, "What Would a Scientific Economics Look Like?", *real-world economics review*, issue no. 47, 3 October 2008, pp. 166-172, <http://www.paecon.net/PAERreview/issue47/Dorman47.pdf>