

January 2005

The Trouble with Mathematics and Statistics in Economics

Deirdre McCloskey

Departments of Economics and History, University of Illinois at Chicago

Department of Philosophy, Erasmus University of Rotterdam

deirdre2@uic.edu

*Il me semble que le moment serait venu
de bien fixer le point si l'on ne veut voir l'économie politique mathématique
s'égarer en toutes sortes de fantaisies stériles qui la déconsidéreront.*

[It seems to me that the moment has come
to clarify this point so that mathematical economics
will not wander off on all kinds of sterile fantasies that will discredit it.]

(Walras to Borkiewicz, Oct 17, 1889,
in Jaffé 1965, p. 384, quoted in Marchionnatti 2004, p. 3)

If you will forgive me, I will be very simple here. The two points I am making are simple, and do not need to be dressed up in fancy clothing. I want to state the points in a way that they cannot be evaded. I'm not optimistic that this tactic will work, but I owe it to a field I love to try. If I am right in my criticism of economics---I pray that I am *not* ---then much of what economists do nowadays is a waste of time. If this is so it is very desirable that we economists do something about it, right now.

I was told recently by a former editor of the *American Economic Review* that he "basically agrees" with what I am saying here.¹ But then he went on to excuse

¹ He had read McCloskey, *The Secret Sins of Economics* Prickley Paradigm Press (University of Chicago Press), 2002, from which much of this paper is taken.

continuing to waste scientific time---for example, in the pages of the *American Economic Review*--- as necessary for the careers of young people. This will not do. Either you agree with me, and will then of course join me in demanding that economics change right away; or you disagree, in which case it is incumbent upon you as a serious scientist to explain exactly why I am mistaken.

The usual objections to mathematical and statistical reasoning in economics seem to me to be unsound. For example, it is often said that economic data is not "strong enough to bear the weight of elaborate mathematics and statistics." People who say this seem to believe that data in, say, physics is superior to that in the social sciences. But a moment's reflection suggests that this cannot in general be true. Data on stock-market transactions, for example, are available in unlimited amounts, whereas data on neutron stars or the early days of universe are strictly limited. Interestingly, it has also been long argued (I find explicit statements in the 1920s, and by Gerard Debreu more recently) that the weakness of the data is a *reason* for mathematics, the notion being that economists must therefore rely on axiom and proof. For the same reason as its opposite, and for some additional reasons, the argument is unsound.

One hears sometimes that economic *theory* is not sufficiently developed to "bear the weight." The intricacies of pure rationality in game theory belie this objection. Surely the assumptions of Nash equilibrium can bear all the mathematical weight one wishes to pile onto them. Some say that mathematics is inherently too "abstract." But that, after all, is the point of any argument, i.e., to abstract from particular circumstances enough to say something interestingly general. That variables other than the life cycle contribute to savings is true, but that is no argument against Franco Modigliani's life and work. And some say, as Walras characterized the position in 1874, that "human liberty

will *never* allow itself to be cast into equations." As Walras replied, "as to those economists who do not know any mathematics, who do not even know what it meant by mathematics and yet have taken the stand that mathematics cannot possibly serve to elucidate economic principles, let them go their way."²

The serious objections are, I think, two. (1.) The kind of mathematics used in economics is typically that of the Department of Mathematics, not that of the departments of Physics or of Engineering. It is existence-theorem, *qualitative* mathematics. It is of no use for science. (2.) The kind of statistics used in economics is that of Department of Statistics, which is *also* a species of "existence theorems." Tests of statistical significance claim to tell *whether* there "exists" an effect of interest rates on investment. But this, too, is of no use for science.

The first, mathematical error has characterized economics since its beginning. It has nothing, really, to do with mathematics, since it can be committed, and was, in entirely verbal economics, such as that of Ricardo. But the coming of Mathematics-Department mathematics, which never asks *how large* something is, has continued this unhappy tradition. As Roberto Marchionnatti notes, the first generations of mathematical economists "were chiefly interested in the problems connected with the relationship between mathematical expressions and experimental [I think he means "experiential"] reality." They "generally seemed not to be worried about the formal establishment of equilibrium, an issue that dominated mathematical economics later."³ He notes that "In the 1930s . . . the axiomatization of economic theory permitted

² Walras 1974/1900 (1954), *Elements of Pure Economics* trans W. Jaffé, p. 47.

³ Roberto Marchionnatti 2004, "On the Application of Mathematics to Political Economy," p. 17).

mathematical developments that were free from problems of the realism of the model used." That's right.

The second, statistical error has come to dominate economics since the cheapening of calculation in the 1970s. It was brought into economics by Tinbergen and Klein in the 1940s, but in statistics generally it dates back to R. A. Fisher in the 1920s and to Karl Pearson in the 1900s.

* * * *

The English political arithmeticians William Petty and Gregory King and the rest in the late seventeenth century---anticipated in the early seventeenth century by, like so much of what we call "English," certain Dutchmen---wanted to know How Much. It was an entirely novel obsession, and perfectly consistent with the Scientific Revolution going on at the time. You might call it bourgeois. How Much will it cost to drain the Somerset Levels? How Much does England's treasure by foreign trade depend on possessing colonies? How Much is this and How Much that? Adam Smith a century later kept wondering how much wages in Edinburgh differed from those in London (too much) and how much the colonies by then acquired in England's incessant eighteenth-century wars against France were worth to the home country (not much). By the late eighteenth century, it is surprising to note, the statistical chart had been invented. What is "surprising" is that it hadn't been invented before--another sign that quantitative thinking was novel, at least in the West (the Chinese had been collecting statistics on population and prices for centuries). European states from Sweden to Naples began in the eighteenth century collecting statistics to worry about: prices, population, balances of trade, flows of gold. The word "statistics" was a coinage of German and Italian

enthusiasts for state action in the early eighteenth century, pointing to a story of the state use of numbering. Then dawned the age of statistics, and everything from drug incarcerations and smoking deaths to the value of a life and the credit rating of Jane Q. Public are numbered.

The formal and mathematical theory of statistics was largely invented in the 1880s by eugenicists---those clever racists at the origin of so much in the social sciences--and perfected in the twentieth century by agronomists at places like the Rothamsted agricultural experiment station in England or at Iowa State University. The newly mathematized statistics became a fetish in fields that wanted to be sciences. During the 1920s, when sociology was a young science, quantification was a way of claiming status, as it became also in economics, fresh from putting aside its old name of political economy, and in psychology, fresh from a separation from philosophy. In the 1920s and 1930s even the social anthropologists counted coconuts.

Mathematics of course is not identical to counting or statistics. There have been some famously good calculators among mathematicians, Leonhard Euler being an instance---he also knew the entire Aeneid by heart; in Latin, I need hardly add. But most of mathematics has nothing to do with actual numbers. Euler used calculation in the same way that mathematicians nowadays use computers, for back-of-the-envelope tests of hunches on the way to developing what the mathematicians are pleased to call a Real Proof of such amazing facts as: $e^{i\pi} + 1 = 0$ (and therefore God exists). You can have a "real" proof, the style of demonstration developed by the Greeks, without examining a single number or even a single concrete example. Thus: the Pythagorean Theorem is true for any right triangle, regardless of its dimensions, and is proven not by induction from many or even millions of numerical examples of right triangles, but universally

and for all time, praise God, may her name be glorified, by deduction from premises. Accept the premises and you have accepted the Theorem. *Quod erat demonstrandum*.

Statistics or other quantitative methods in science (such as accounting or experiment or simulation) answer inductively How Much. Mathematics by contrast answers deductively Why, and in a refined and philosophical version very popular among mathematicians since the early nineteenth century, Whether. "Why does a stone dropped from a tower go faster and faster?" Well, $F = ma$, understand? "I wonder Whether the mass, m , of the stone has any effect at all." Well, yes, it does: notice that there's a little m in the answer to the Why question.

Why/Whether is not the same question as How Much. You can know that forgetting your lover's birthday will have *some* effect on your relationship (Whether), and even understand that the neglect works through such-and-such an understandable psychological mechanism ("Don't you love me enough to know I care about birthdays?" Why). But to know How Much the neglect will hurt the relationship you need to have in effect numbers, those *ms* and *as*, so to speak, *and some notion of their magnitudes*. Even if you know the Why (the proper theory of the channels through which forgetting a birthday will work; again by analogy, $F = ma$), the How Much will depend on exactly, numerically, quantitatively how sensitive this or that part of the Why is in fact in your actual beloved's soul---how much in this case the m and a are. And such sensitivity in an actual world, the scientists are always saying, is an empirical question, not theoretical. "All right, you louse, that's the last straw: I'm moving out" or "Don't worry, dear: I know you love me" differ in the sensitivity, the How Much, the quantitative effect, the magnitude, the mass, the oomph.

Economics since its beginning has been very often "mathematical" in this sense of being interested in Why/Whether arguments without regard to How Much. For example: If you buy a loaf of bread from the supermarket both you and the supermarket (its shareholders, its employees, its bread suppliers) are made to some degree better off. Economists have long been in love with this simple argument. They have since the eighteenth century taken the argument a crucial and dramatic step further: that is, they have deduced something from it, namely, Free trade is neat. If each deal between you and the supermarket, and the supermarket and Smith, and Smith and Jones, and so forth is betterment-producing (a little or a lot: we're not talking quantities here), then (note the "then": we're talking about deduction here) free trade between the entire body of Italian people and the entire body of English people is betterment producing, too. And therefore (note the "therefore") free trade between any two groups is neat. The economist notes that if all trades are voluntary they all have some gain. So free trade in all its forms is neat. For example, a law restricting who can get into the pharmacy business is a bad idea, not neat at all, because free trade is good, so non-free trade is bad. Protection of French workers is bad, because free trade is good. And so forth, to literally thousands of policy conclusions.

Though it is among the three or four most important arguments in economics, it is not empirical. It contains no statements of How Much. It says there *exists* a gain from trade. "I wonder Whether there exists [*in whatever quantity*] a good effect of free trade." Yes, one exists: examine this page of mathematics; look at this diagram; listen to my charming parable about you at the supermarket. Don't ask How Much. The reasoning is Why/Whether.

As stated it cannot be wrong, no more than the Pythagorean Theorem can be. It's not a matter of approximation, not a matter of How Much. It's a chain of logic from implicit axioms (which can be and have been made explicit, in all their infinite variety) to a "rigorous" qualitative conclusion (in it's infinite variety). Under such-and-such a set of assumptions, A, the conclusion, C, must be that people are made better off. A implies C, so free trade is beneficial anywhere.

Philosophers call this sort of thing "valid" reasoning, by which they do not mean "true," but "following from the axioms---if you believe the axioms, such as A, then C also must be true." If you believe that any individual exchange arrived at voluntarily is good, then with a few extra assumptions (e.g., about the meaning of "voluntarily"; or, e.g., about how one person's good depends on another's) you can get the conclusion that free international trade among nations is good.

Why/Whether reasoning, which is also characteristic of the Department of Mathematics, could be called philosophical. The Department of Philosophy has a similar fascination with deduction, and a corresponding boredom with induction. Neither Department bothers with How Much. In the Philosophy Department either relativism is or is not open to a refutation from self-contradiction. It's not a *little* refuted. It's knocked down, or not. In the Department of Mathematics the Goldbach Conjecture, that every even number is the sum of two prime numbers (e.g., $24 = 13 + 11$), is either true or false---or, to introduce a third possibility admitted since the 1930s, undecidable. Supposing it's decidable, there's no question of How Much. You can't in the realm of Why/Whether, in the Department of Mathematics or the Department of Philosophy or some parts of the Department of Economics, be a little bit pregnant.

Since about 1947 the front line and later the dominant and by now the arrogantly self-satisfied and haughtily intolerant if remarkably unproductive scientific program in economics has been to reformulate verbal---but still philosophical/mathematical, i.e. qualitative, i.e. Why/Whether---arguments into symbols and variables and diagrams and fixed point theorems and the like. The program is called "Samuelsonian." Paul Samuelson and his brother-in-law Kenneth Arrow led the movement to be explicit about the math in economics, against great opposition. They were courageous pioneers. In 1947 Samuelson set the tone with the publication of his Ph. D. dissertation, which had been finished in 1941. In 1951 Arrow carried it to still higher realms of Department-of-Mathematics mathematics with his own Ph.D. dissertation. Their enemies, a few of whom are still around, said, with the humanists, " This math stuff is too hard, too inhuman. Give me words. Sentiment. Show me some verbal argumentation or some verbal history. Or even actual numbers. But none of this new x and y stuff. It gives me a headache."

But there was nothing whatever new about deductive reasoning in economics in 1947. In the 1740s and 1750s David Hume in Scotland and the physiocrats in France were busy inventing philosophical, entirely qualitative, Why/Whether arguments about economics). Deducing sometimes surprising and anyway logically valid (if not always true) conclusions from assumptions about the economy is a game economists have always loved. If you want to connect one thing with another, deduce conclusions C from assumptions A, free trade from characterizations of an autonomous consumer, why not do it universally and for all time? Why not, asked Samuelson and Arrow and the rest, with much justice, do it right? Getting deductions right is the Lord's work, if

not the only work the Lord favors. Like all virtues it can be carried too far, and be unbalanced with other virtues, becoming the Devil's work, sin. But all virtues are like that.

True, for practical purposes of surveying grain fields it would work just as well as Pythagoras' Greek proof to have a Babylonian-style of proof-by-calculation showing that the sums of squares of the sides of millions of triangles seem to be pretty much equal to the sums of squares of their hypotenuses. You might make a similar case for the free trade theorem, noting for example that the great internal free-trade zone called the United States still has a much higher average income (20 to 30 percent higher) than otherwise clever and hard working countries like Japan or Germany, which insist on many more restrictions on internal trade, such as protection of small retailing. And, true, the improvement of computers is making more Babylonian-style "brute force calculations" (as the mathematicians call them with distaste) cheaper than some elegant formulas ("analytic solutions," they say, rapturously). Economics, like many other fields--architecture, engineering---is about to be revolutionized by computation.

But if beyond clumsy fact or numerical approximation there is an elegant and exact formula--- $F = ma$ or $E = mc^2$ or, to give a somewhat less elegant example from economics, $1 + i_{usa} = (e_{forward} / e_{spot}) (1 + i_{france})$, "covered interest arbitrage"---why not use it? Of course, any deduction depends on the validity of the premises. If a sufficiently high percentage of potential arbitrageurs in the markets for French and U.S. bonds and currency are slothful, then covered interest arbitrage will not hold. But likewise any *induction* depends on the validity of the data. If the sample used to

test the efficacy of mammograms in preventing premature death is biased, then the statistical conclusions will not hold. Any calculation depends on the validity of the inputs and assumptions.

* * * *

A real science, or any intelligent inquiry into the world, whether the study of earthquakes or the study of poetry, economics or physics, history or anthropology, art history or organic chemistry, a systematic inquiry into one's lover or a systematic inquiry into the Italian language, must do two things. If it only does one of them it is not an inquiry into the world. It may be good in some other way, but not in the double way that we associate with good science or other good inquiries into the world, such as a detective solving a case.

I am sure you will agree: An inquiry into the world must think and it must look. It must theorize and must observe. Formalize and record. Both. That's obvious and elementary. Not everyone involved in a collective intelligent inquiry into the world need do both: the detective can assign his dim-witted assistant to just observe. But the inquiry as a whole must reflect and must listen. Both. Of course.

Pure thinking, such as mathematics or philosophy, is not, however, to be disdained, not at all. Euler's equation, $e^{i\pi} + 1 = 0$, really is quite remarkable, linking "the five most important constants in the whole of analysis" (as Philip Davis and Reuben Hersh note), and would be a remarkable cultural achievement even if it had no worldly use. But certainly the equation is not a result of looking at the world. So it is not science; it is a kind of abstract art. Mathematicians are proud of the uselessness of most of what

they do, as well they might be: Mozart is "useless," too; to what would you "apply" the Piano Sonata in A?

Nor is pure, untheorized observation to be disdained. There is something in narration, for example, that is untheorizable (though it is surprising to non-humanists how much of it can and has recently been theorized by literary critics). At some level a story is just a story, and artful choice of detail within the story is sheer observation-not brute observation, which is a hopeless ambition to record everything, but sheer.

So pure mathematics, pure philosophy, the pure writing of pure fictions, the pure painting of pictures, the pure composing of sonatas are all, when done well or at least interestingly, admirable activities. I have to keep saying "pure" because of course it is entirely possible-indeed commonplace for novelists, say, to take a scientific view of their subjects (Balzac, Zola, Sinclair Lewis, the post-War Italian realists, among many others are well known for their self-conscious practice of a scientific literature; Roman satire is another case; or Golden Age Dutch painting). Likewise scientists use elements of pure narration (in evolutionary biology and economic history) or elements of pure mathematics (in physics and economics) to make scientific arguments.

I do not want to get entangled in the apparently hopeless task of solving what is known as the Demarcation Problem, discerning a line between science and other activities. It is doubtful such a line exists. The efforts of many intelligent philosophers of science appear to have gotten exactly nowhere in solving it. I am merely suggesting that a science like many other human practices such as knitting or making a friend should be about the world, which means it should attend to the world. And it should also be something other than miscellaneous facts, such as the classification of animals in the Chinese *Celestial Emporium of Benevolent Knowledge* noted by Borges: (a.) those that

belong to the Emperor, (b.) embalmed ones, (c.) those that are trained, (d.) suckling pigs, (e.) mermaids, and so forth, down to (n.) those that resemble flies from a distance. Not brute facts. And not mere theory.

So I am not dragging economics over to some implausible definition of Science and then convicting it of not corresponding to the definition. Such a move is common in economic methodology-for example in some of the less persuasive writings of the very persuasive economist Marc Blaug. I am merely saying that economists want to be involved in an intelligent inquiry into the world. If so, the field as a whole must theorize and observe, both. This is not controversial.

An economist at a leading graduate program listening to me will now burst out with: "Great! I entirely agree: theorize and observe, though of course as you admit we can specialize in one or the other as long as the whole field does both. And that, Deirdre, is exactly what we already do, on a massive scale. And we do it very well, if I don't say so myself. We do very sophisticated mathematical theorizing, such as in the Mas-Collel, Whinston, and Green textbook (1995), and then we test the theory in the world using very tricky econometrics, such as Jeffrey M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data* (2001). You can see the results in any journal of economics. Some of it is pure theory, some econometrics. Theorize and observe."

* * * *

To which I say: Rubbish. She and her colleagues, when they are being most highbrow and Science-proud, don't really do either theorizing or observing. Economics in its most prestigious and academically published versions engages in two activities, qualitative theorems and statistical significance, which look like theorizing and

observing, and have (apparently) the same tough math and tough statistics that actual theorizing and actual observing would have. But neither of them is what it claims to be. Qualitative theorems are not theorizing in a sense that would have to do with a double-virtued inquiry into the world. In the same sense, statistical significance is not observing.

It is not difficult to explain to *outsiders* what is so dramatically, insanely, sinfully wrong with the two leading methods in high-level economics, qualitative theorems and statistical significance. It is very difficult to explain it to insiders, because the insiders cannot believe that methods in which they have been elaborately trained and which are used by the people they admire most are simply unscientific nonsense, having literally nothing to do with whatever actual scientific contribution (and I repeat, it is considerable) that economics makes to the understanding of society. So they simply can't grasp arguments that are plain to people not socialized in economics.⁴

Why-Whether reasoning in economics takes this form: A implies C. The crucial point is that the A and the C are indeed qualitative. They are not of the form "A is 4.8798." They are of the qualitative form, "A is 'everyone is motivated by P-Only considerations'," say, which implies "free trade is neat." No numbers. You realize your lover will be annoyed by the neglected birthday to some degree, but we're not talking about magnitudes. Why/Whether. Not How Much. The economic "theorists" focus on existence theorems. With such and such general (or not so general, but anyway non-quantitative) assumptions A there exists a state of the imagined world C. A typical statement in economic "theory" is, "if information is symmetric, an equilibrium of the

⁴ See Chapters 10-13 in *Knowledge and Persuasion in Economics* [1994] and Chapters 7 and 8 in *The Rhetoric of Economics* [2nd ed. 1998].

game exists" or, "if people are rational in their expectations in the following sense, buzz, buzz, buzz, then there exists an equilibrium of the economy in which government policy is useless."

For example, the non-free traders, often European and disproportionately nowadays French, point out that you can make other assumptions about how trade works, A', and get other conclusions, C' not so favorable to laissez faire. The free-trade theorem, which sounds so grand, is actually very easy to overturn, and numerous careers have been built in economics doing so (Paul Krugman's, for example). Suppose a big part of the economy---say the household---is, as the economists say, "distorted" (e.g., suppose people in households do things for love: you can see that we economists have a somewhat peculiar idea of "distortion"). Then it follow rigorously (that is to say, mathematically) that free trade in other sectors (e.g. manufacturing) will not be the best thing. In fact it can make the average person worse off than restricted, protected, tariffed trade would.

The theorists don't have to operate in this existence-theorem way. They could instead---some do---use mathematics to develop functional forms into which the world's data can be plugged. It is the difference between abstract general equilibrium---a field of economics which practically everyone now agrees was a complete waste of time and talent---and computable general equilibrium, which has nothing whatever to do with the existence theorems and has everything to do with picking sensible numbers and simulating.

The trouble with the qualitative theorems recommended by Paul Samuelson in the *Foundations* is this. Naturally, if you change assumptions (introducing households who do not operate on Prudence-Only motivations, say; or [I speak now to insiders]

making information a little asymmetric; or [ditto] introduce any Second Best, such as monopoly or taxation; or [ditto] nonconvexities in production) in general a conclusion about free trade is going to change. There's nothing deep or surprising about this: changing your assumptions changes your conclusions. Call the new conclusion C' . So we have the old A implies C and the fresh, publishable novelty, A' implies C' . But, as the mathematicians say, we can add another prime and proceed as before, introducing some other plausible possibility for the assumptions, A'' (read it "A double prime"), which implies its own C'' . And so forth: A''' implies C''' . And on and on and on and on, until the economists get tired and go home.

What has been gained by all this? It is pure thinking, philosophy. It is not disciplined by any simultaneous inquiry into How Much. It's qualitative, not quantitative, and not organized to allow quantities into the story. It's like stopping with the conclusion that forgetting your lover's birthday will have some bad effect on one's relationship---you still have no idea how much, whether trivial or disastrous or somewhere in between. So the pure thinking is unbounded. It's a game of imagining how your lover will react endlessly. True, if you had good ideas about what were plausible assumptions to make, derived from some inquiry into the actual state of the world, the situation might be rescued for science and other inquiries into the world, such as the inquiry into the probably quantitative effect of missing a birthday on your lover's future commitment to you. But if not---and such is the usual practice of "theoretical" pieces in economics, about half the items in any self-respecting journal of economic science---it's "just" an intellectual game.

I have expressed admiration for pure mathematics and for Mozart's concertos.

Fine. But economics is supposed to be an inquiry into the world, not pure thinking. (If it is to be justified as pure thinking, just "fun," it is not very entertaining. No one would buy tickets to listen to a "theory" seminar in economics. As mathematical entertainment the stuff is quite poor.) The A-prime/C-prime, existence-theorem, qualitative-only "work" that economists do is like chess problems. Chess problems usually do not have anything to do even with playing real chess---since the situations are often ones that could not arise in a real game. And chess itself has nothing to do with living, except for its no doubt wonderful purity as thought, *à la* Mozart.

What kind of theory would actually contribute to a double-virtued inquiry into the world? Obviously, it would be the kind of theory for which actual numbers can conceivably be assigned. If Force equals Mass times Acceleration you have a potentially quantitative insight into the flight of cannon balls, say. But the qualitative theorems don't have any place for actual numbers. So the "results" keep flip-flopping, endlessly, pointlessly.

Samuelson himself famously showed in the 1940s that "factor prices" (such as wages) are "equalized" by trade in steel and wheat and so forth-as a qualitative theorem, under such and such assumptions, A. It could be an argument against free trade. But shortly afterwards it was shown by Samuelson himself, among others that if you make alternative assumptions, A', you get very different conclusions. And so it went, and goes, with the limit achieved only in boredom, all over economics. Make thus-and-such assumptions, A, about the following game-theoretic model and you can show that a group of unsocialized individuals will form a civil society. Make another set of assumptions, A', and they won't. And so on and so forth. Blah, blah, blah, blah, to no scientific end.

Such stuff has taken over fields near to economics, first political science and now increasingly sociology. A typical "theoretical" paper in the *American Political Science Review* shows that under assumptions A the comity of nations is broken; in the next issue someone will show that under A' it is preserved. This is not theory in the sense that, say, physics uses the term. Pick up a copy of the *Physical Review* (it comes in four versions; pick any). Open it at random. You will find mind-breakingly difficult mathematics, and physics that no one except a specialist in the particular tiny field can follow. But always, on every page, you will find repeated, persistent attempts to answer the question How Much. Go ahead: do it. Don't worry; it doesn't matter that you can't understand the physics. You will see that the physicists use in nearly every paragraph a rhetoric of How Much. Even the theorists as against the experimenters in physics spend their days trying to figure out ways of calculating magnitudes. The giveaway that something other than scientific is going on in "theoretical" economics (and, alas, political science) is that it contains not, from beginning to end of the article, a single attempt at a magnitude.

* * * *

"But wait a minute, Deirdre," the Insider Economist breaks in (he is getting very, very annoyed because, as I told you, he Just Doesn't Get It). "You admitted that we economists also do econometrics, that is, formal testing of economic hypotheses using advanced statistical theory. You, as an economist, can hardly object to specialization: some people do theory, some empirical work."

Yes, my dear young colleague. Since I have been to your house and noted that you have not a single work on economics before your own graduate training I suppose

you are not aware that the argument was first made explicit in 1957 by Tjalling Koopmans, a Dutch-American economist at Yale (Nobel 1975), who in his *Three Essays on the State of Economic Science* recommended just such a specialization. He recommended that "theorists" spend their time on gathering a "card file" of qualitative theorems attaching a sequence of axioms A^I, A^{II}, A^{III} , etc. to a sequence of conclusions C^I, C^{II}, C^{III} , etc., separated from the empirical work, "for the protection [note the word, students of free trade] of both."

Now this would be fine if the theorems were not qualitative. If they took the form that theorems do in physics (better called "derivations," since physicists are completely uninterested in the existence theorems that obsess mathematicians and philosophers), good. Then the duller wits like Deirdre McCloskey the economic historian could be assigned to mere observation, filling in blanks in the theory. But there are no blanks to fill in, no How Much questions asked, in the theory that economists admire the most and that has taken over half of their waking hours.

Still, things would not be so bad if on the lower-status empirical side of academic economics all was well. The empiricists like me in their dull-witted way could cobble together actual scientific hypotheses, simply ignoring the "work" of the qualitative theorists. Actual players of chess could ignore the "results" from chess problems. In effect this is what happens. The "theories" proffered by the "theorists" are not tested. In their stead linearized models that try crudely to control for this or that effect are used. An empiricist could therefore try to extract the world's information about the price sensitivity of demand for housing in Britain in the 1950s, say.

But unhappily the empirical economists *also* have become confused by qualitative "results." They, too, have turned away from one of the two questions

necessary for a serious inquiry into the world (the other is Why), How Much. The sin sounds improbable, since empirical economics is drenched in numbers, but the numbers they acquire with their most sophisticated tools---as against their most common tools, such as simple enumeration and systems of accounting---are it turns out meaningless.

The confusion and meaninglessness arises from "statistical significance." It has become since the cheapening of computation in the 1970s a plague in economics, in psychology, and, most alarmingly, in medical science. Consider the decades-long dispute over the prescribing of routine mammograms to screen for early forms of breast cancer. One school says, Start at age 40. The other says, No, age 50. (And still another, Never routinely. But set that aside.) Why do they differ? The American nurses' epidemiological study or the Swedish studies on which the empirical arguments are based are quite large. But there's a lot of noise in the data. So: although starting as early as age 40 does seem to have some effect, the samples are not large enough to be conclusive. By what standard? By the standard of *statistical* significance at the 5%, 1%, 0.1%, or whatever level.

So the situation is this. The over-50 school admits that there is some positive effect in detecting early cancers from starting mammograms as early as age 40; but, they say with a sneer, it's uncertain. You'll be taking some chance of being fooled by chance. Nasty business. Really, something to avoid. Even though there is a life-saving effect of early mammograms in the data on average, Mr. Medical Statistician is uncomfortable about claiming it. The purpose of medical research is to save lives. His comfort is not what we are chiefly concerned with. The data is noisy. It is a pity that God arranged it

that way. She should have been more considerate. But She's done what She's done. Now we have to decide if the cost of the test is worth the benefit. And the data shows a benefit.

Mr. Medical Statistician replies, with some indignation: "No it's does not. At conventional levels of significance there is no effect."

Deirdre, with more indignation: "Nonsense. You are trying, alas, to make a qualitative judgment of existence. Compare the poor, benighted Samuelsonian "theorist." We always in science need How Much, not Whether. The effect is empirically there, whatever the noise is. If someone called "Help, help!" in a faint voice, in the midst of lots of noise, so that at the 1% level of significance (the satisfactorily low probability that you will be embarrassed by a false alarm) it could be that she's saying "Kelp, kelp!" (which arose perhaps because she was in a heated argument about a word proposed in a game of Scrabble), you wouldn't go to her rescue?

The relevant and quantitative question about routine mammograms, which has recently been reopened, is the balance of cost and benefit, since there could be costs (such as deaths from intrusive tests resulting from false positives) that offset the admittedly slight gain from starting as early as age 40. But suppose, as was long believed, that the costs do not offset the gain. That the net gain is slight is no comfort to the (few) people who die unnecessarily at 42 or 49 on account of Mr. Medical Statistician's gross misunderstanding of the proper role of statistics in scientific inquiries. A death is a death. The over-50 people are killing patients. Maybe only slightly more than zero patients. But more than zero is murder.

Or consider the aspirin-and-heart-attack studies. Researchers were testing

the effects of administering half an aspirin a day to men who had already suffered a heart attack. To do the experiment correctly they gave one group the aspirin and the other a placebo. But they soon discovered—well short of conventional levels of statistical significance—that the aspirin reduced reoccurrences of heart attacks by about a third. What did they do? Did they go on with the study until they got a large enough sample of dead placebo-getters to be sure of their finding at levels of statistical significance that would make the referees of cardiology journals happy? Of course not: that would have been shockingly (though not unprecedentedly) unethical. They stopped the study, and gave everyone aspirin. (A *New Yorker* cartoon around the same time made the point, showing a tombstone inscribed, "John Smith, Member, Placebo Group.")

Or consider public opinion polls about who is going to win the next presidential election. These always come hedged about with warnings that the "margin of error is 2% plus or minus." So is the claim that prediction of a presidential election six months before it happens is only 2% off? Can that be reasonable? What is being reported is the sampling error (and only at conventional levels of significance, themselves arbitrary). An error caused, say, by the revelation two months down the road that one of the candidates is an active child molester is not reckoned as part of "the error." You can see that a game is being performed here. The statement of a "probable error" of 2% is silly. A tiny part of all the errors that can afflict a prediction of a far-off political event is being elevated to the rhetorical status of The Error. "My streetlight under sampling theory is very bright, so let's search for the keys under the streetlight, even though I lost them in the dark." This cannot go on.

The point here is that such silliness utterly dominates empirical economics.

In a study of all the empirical articles in the *American Economic Review* in the 1980s it was discovered that fully 96% of them confused statistical and substantive significance.⁵ A follow-up study for the *AER* in the 1990s found that the problem had become worse.⁶

The problem is that a number fitted from the world's experiments can be important economically without being noise-free. And it can be wonderfully noise-free without being important. On the one hand: It's obvious, you will agree, that a "statistically insignificant" number can be very significant for some human purpose. If you really, truly want to know how the North American Free Trade Agreement affected the average worker in the United States, then it's too bad if the data are noisy, but that's not the point. You really, truly want to know it. You have to go with what God has provided. And on the other hand: It is also obvious that a "statistically significant" result can be insignificant for any human purpose. When you are trying to explain the rise and fall of the stock market it may be that the fit is very tight for some crazy variable, say skirt lengths (for a long while the correlation was actually quite good). But it doesn't matter: the variable is obviously crazy. Who cares how closely it fits? For a long time in Britain the number of ham radio operator licenses granted annually was very highly correlated with the number of people certified insane. Very funny.

In short, statistical significance is neither necessary nor sufficient for a result to be scientifically significant. Most of the time it is irrelevant. A researcher is simply committing a scientific error to use it as it is used in economics and the other social

⁵ Deirdre McCloskey and Stephen Ziliak, "The Standard Error of Regression," *Journal of Economic Literature*, March 1996

⁶ Stephen Ziliak and Deirdre McCloskey, "Size Matters: The Standard Error of Regressions in the *American Economic Review* during the 1990s." Subject of a symposium in the *Journal of Socio-Economics*, Autumn 2004, with comments by Kenneth Arrow, Arnold Zellner, Clive Granger, Edward Leamer, Joel Horowitz, Erik Thorbecke, and others.

sciences and in medical science and (a strange one, this) population biology as an all-purpose way of judging whether a number is large enough to matter. Mattering is a human matter; the numbers figure, but after collecting them the mattering has to be decided finally by us; mattering does not inhere in a number.

Physics and chemistry, though of course highly numerical, hardly ever use statistical significance. Economists and those others use it compulsively, mechanically, erroneously to provide a non-controversial way of deciding whether or not a number is large. You can't do it this way. No competent statistical theorist has disagreed with me on this point since Neyman and Pearson in 1933. There is no mechanical procedure that can take over the last, crucial step of an inquiry into the world, asking How Much in human terms that matter.

* * * *

The argument is not against statistics in empirical work, no more than it is against mathematics in theoretical work. It is against certain very particular and peculiar practices of economic science and a few other fields. Economics has fallen for qualitative "results" in "theory" and significant/insignificant "results" in "empirical work." You can see the similarity between the two. Both are looking for on/off findings that do not require any tiresome inquiry into How Much, how big is big, what is an important variable, how much exactly is its oomph. Both are looking for machines to produce publishable articles. In this last they have succeeded since Samuelson spoke out loud and bold beyond the dreams of intellectual avarice. Bad science-using qualitative theorems with no quantitative oomph and statistical significance also with no quantitative oomph-has driven out good.

The progress of economic science has been seriously damaged. Most of what appears in the best journals of economics is therefore mistaken. I find this unspeakably sad. All my friends, my dear, dear friends in economics, have been wasting their time. You can see why I am agitated about the Two Sins. They are vigorous, difficult, demanding activities, like hard chess problems. But they are worthless as science.

The physicist Richard Feynman called such activities Cargo Cult Science. Certain New Guinea tribesmen had prospered mightily during the Second World War when the American Air Force disgorged its cargo to fight the Japanese. After the War the tribesmen wanted the prosperity to come back. So they started a "cargo cult." Out of local materials they built mock airports and mock transport planes. They did an amazingly good job: the cargo-cult airports really do look like airports, the planes like planes. The only trouble is, they aren't actually. Feynman called sciences he didn't like "cargo cult sciences. By "cargo cult" he meant that they looked like science, had all that hard math and statistics, plenty of long words; but actual science, actual inquiry into the world, was not going on.

I am afraid that my science of economics has come to the same point. Paul Samuelson, though a splendid man and a wonderful economist (honestly), is a symbol of the pointlessness of qualitative theorems. Samuelson, actually, is more than merely a symbol-he made and taught and defended the Two Sins, at one time almost singlehandedly. It was a brave stance. But it had terrible outcomes. Samuelson advocated the "scientific" program of producing qualitative theorems, developing qualitative-theorem-generating-functions, such as "revealed preference" and "overlapping generations" models" and above all the machinery of Max U. He was involved also, it turns out somewhat surprisingly, in the early propagation of

significance testing, the "scientific" method of empirical work running on statistical significance without a loss function, through his first Ph. D. student, Lawrence Klein. So it is only fair to call both the sins of modern economics Samuelsonian.

Until economics stops believing, contrary to its own principles, that an intellectual free lunch is to be gotten from qualitative theorems and statistical significance it will be stuck on the ground waiting at the cargo-cult airport, at any rate in its high-end activities uninterested in (Really) How Much. High-end theoretical and econometric papers will be published. Careers will be made, thank you very much. Many outstanding fellows (and no women) will get chairs at Princeton and Chicago. But our understanding of the economic world will continue to stagnate.

Appendix:

Some Earlier Attempts to Make These Points, Aside from the Items Mentioned in the Footnotes

- "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review*, Supplement 75 (2, May 1985): 201-205.
- "Formalism in Economics, Rhetorically Speaking," *Ricerche Economiche* 43 (1989), 1-2 (Jan-June): 57-75. Reprinted with minor revisions in *American Sociologist* 21 (1, Spring, 1990): 3-19.
- "Economic Science: A Search Through the Hyperspace of Assumptions?" *Methodus* 3 (1, June 1991): 6-16.
- Knowledge and Persuasion in Economics*. Cambridge University Press 1994.
- The Vices of Economists; The Virtues of the Bourgeoisie*. University of Amsterdam Press and University of Michigan Press, 1997. Japanese translation, with new preface for Japanese readers by McCloskey, Tokyo: Chikuma Shobo, Ltd., 2002
- How to Be Human* *Though an Economist*. Ann Arbor: University of Michigan Press, 2000.
- [with Stephen Ziliak] *Size Matters: How Some Sciences Lost Interest in Magnitude, And What to Do About It* } Full outline available; to be drafted by May, 2005.