Randall Morck and Bernard Yeung

Economics, History, and Causation

Economics and history both strive to understand causation: economics by using instrumental variables econometrics, and history by weighing the plausibility of alternative narratives. Instrumental variables can lose value with repeated use because of an econometric tragedy of the commons: each successful use of an instrument creates an additional latent variable problem for all other uses of that instrument. Economists should therefore consider historians' approach to inferring causality from detailed context, the plausibility of alternative narratives, external consistency, and recognition that free will makes human decisions intrinsically exogenous.

Economics and history have not always got on. Edward Lazear's advice that all social scientists adopt economists' toolkit evoked a certain skepticism, for mainstream economics repeatedly misses major events, notably stock market crashes, and rhetoric can be mathematical as easily as verbal.[1] Written by winners, biased by implicit assumptions, and innately subjective, history can also be debunked.[2] Fortunately, each is learning to appreciate the other. Business historians increasingly use tools from mainstream economic theory, and economists display increasing respect for the methods of mainstream historians.[3] Each

Partial funding from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged by Randall Morck.

[1] Edward Lazear, "Economic Imperialism," *Quarterly Journal of Economics* 116, no. 1 (Feb. 2000): 99–146; Irving Fischer, "Statistics in the Service of Economics," *Journal of the American Statistical Association* 28, no. 181 (Mar. 1933): 1–13; Deirdre N. McCloskey, *The Rhetoric of Economics* (Madison, Wisc., 1985).

[2] Henry Ford, with Samuel Crowther, *My Life and Work* (Garden City, N.Y., 1922), 43–44; Marshall McLuhan, *The Gutenberg Galaxy: The Making of Typographic Man* (Toronto, 1962); Jacques Derrida, *De la Grammatologie* (Paris, 1967).

[3] Naomi R. Lamoreaux, Daniel M. G. Raff, and Peter Temin, "Economic Theory and Business History," in *The Oxford Handbook of Business History*, ed. Geoffrey Jones and Jonathan Zeitlin (Oxford, 2007), ch. 3; Alfred D. Chandler Jr., *Strategy and Structure: Chapters in the History of the Industrial Enterprise* (Cambridge, Mass., 1962); Mira Wilkins, *The Emergence of Multinational Enterprise* (Cambridge, Mass. 1970); Peter Hertner and Geoffrey

Business History Review 85 (Spring 2011): 39–63. doi:10.1017/S000768051100002X
 ISSN 0007-6805; 2044-768X (Web).

field has infirmities, but also strengths. We propose that their strengths usefully complement each other in untangling the knotty problem of causation.

This complementarity is especially useful to economics, where establishing what causes what is often critical to falsifying a theory. Carl Popper argues that scientific theory advances by successive falsifications, and makes falsifiability the distinction between science and philosophy.[4] Economics is not hard science, but nonetheless gains hugely from a now nearly universal reliance on empirical econometric tests to invalidate theory. Edward O. Wilson puts it more bluntly: "Everyone's theory has validity and is interesting. Scientific theories, however, are fundamentally different. They are designed specifically to be blown apart if proved wrong; and if so destined, the sooner the better."[5] Demonstrably false theories are thus pared away, letting theoreticians focus on as yet unfalsified theories, which include a central paradigm the mainstream of the profession regards as tentatively true.[6] The writ of empiricism is now so broad that younger economists can scarcely imagine a time when rhetorical skill, rather than empirical falsification, decided issues, and the simplest regression was a day's work with pencil and paper.

But such was once the case. Relying on common sense, Thomas Malthus writes, "Population, when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio."[7] Francis Edgeworth, relying on introspection, affirms a gender-specific "capacity for pleasure" and "a nice consiliance between the deductions of the utilitarian principle and the disabilities and privileges which hedge around modern womanhood."[8] John K. Galbraith, relying on a keen intellect, declares that "competitors of General Motors are especially unlikely to initiate price reductions that might provoke further and retributive price cutting. . . . Everyone knows that the survivor of such a contest would not be the aggressor but General Motors."[9] And a little data can be a dangerous thing—for example, cold war–era editions of Paul Samuelson's classic textbook, *Economics*, feature graphs of Soviet GNP surpassing

Jones, *Multinationals: Theory and History* (Aldershot, U.K., 1986); Geoffrey Jones, *Multinationals and Global Capitalism: From the Nineteenth to the Twenty-First Century* (Oxford, 2005); and many others.

[4] Carl Popper, *Logik der Forschung* (Vienna, 1934).

[5] Edward O. Wilson, *Consilience: The Unity of Knowledge* (New York, 1998), 47.

[6] Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1962).

[7] Thomas Malthus, *An Essay on the Principle of Population* (London, 1798), 4.

[8] Francis Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London, 1881), 77, 79.

[9] John Kenneth Galbraith, *The New Industrial State* (London, 1967), 36.

U.S. GNP by the 1980s, or 1990s at the latest, based on simple extrapolations from past trends.[10]

These indisputably great economists wrote as they did because their introspection, common sense, intellects, and observations shaped their thoughts. Rhetorical flourish usefully prevented their economics from lapsing into a treatment for insomnia, but what these old masters did was not science, but something closer to history. For historians, too, weave common sense, introspection, intellect, and historical records into narratives that explain the past and illuminate the present.

Their work added much to economics. Edgeworth, Samuelson, Leon Walras, and Wassily Leontief brought algebraic clarity to elegant narratives spun by Adam Smith, John Stuart Mill, Voltaire, and Karl Marx; and the combination was genuinely powerful. But so are folk tales, like Rudyard Kipling's *"Just So" Stories* (1902), which relate how the camel got his hump, how the leopard got his spots, and so on. Good narratives are compelling, socially edifying, and plausible explanations of why things are "just so." The critical difference is evidence.

This lesson is now so deeply accepted that one seldom sees an economic theory article without valid econometric evidence, or at least a compelling survey of supportive empirical evidence. This is an unmitigated blessing. Empirical observation has pushed extremists toward the center, for the data undermine both Marxism and perfect markets. The twenty-first-century left contemplates a toilet-trained capitalism.[11] The twenty-first-century right frets over entrenched oligarchs, the potential importance of fiscal policy, and the optimal design of government.[12] Frenzied cries to abandon either markets or government can still be heard elsewhere on university campuses, but rarely amid economists. Our debates remain passionate, but are far more clinical and data driven than before computers and mass storage ushered in the Age of Data.

But economics is more than econometrics; it is an ongoing interplay of theory and evidence. Thomas Kuhn argues that science establishes paradigms—structural theories of what causes what—that remain valid as long as they are not inconsistent with extant empirical

[10] David Levy and Sandra Peart, "Soviet Growth and American Textbooks," Economics Department, George Mason University working paper, 2009.

[11] Paul Krugman, *The Conscience of a Liberal* (New York, 2007); Jeffrey Sachs, *The End of Poverty: Economic Possibilities for Our Time* (New York, 2005); Joseph Stiglitz, *Making Globalization Work* (New York, 2006).

[12] Raghuram Rajan and Luigi Zingales, *Saving Capitalism from the Capitalists* (Princeton, 2004); Martin Feldstein, "Rethinking the Role of Fiscal Policy," *American Economic Review* 99, no. 2 (2009): 556–59; James Buchanan, "Public Choice: The Origins and Development of a Research Program," Center for the Study of Public Choice at George Mason University, 2003.

evidence.[13] The overwhelming success of econometrics in fundamentally altering the way economists think and debate attracts attention, and therefore critics. Speaking for many of these, Fischer Black blasts econometrics for confusing "correlation with causation" and econometricians for terminology that propounds that confusion.[14] Black's attack hit hard, and endogeneity bias, previously but one of many potential econometric problems, became "the" econometric problem.

Thus rattled, economists returned to history, searching for tools with which to cultivate better econometrics. An assortment of econometric techniques based on instrumental variables became "the" response to Black's critique. Economists often look to history for instrumental variables: factors determined long ago that cannot possibly "be caused" by things going on today. If paths of causation can be traced through such factors, the direction of causality can be inferred.

This technique is very powerful where it can be applied—for example, in natural experiments.[15] However, econometrically useful natural experiments are few and far between, so economists often make do with iffier instrumental variables techniques. We argue that strict limitations on the validity of instrumental variables greatly limit their utility, and that repeated use of the same instrumental variables in related economic contexts undermines their validity in an econometric tragedy of the commons. However, we believe that economists might find other ways of establishing causality by recognizing history as more than a toolshed for instrumental variables. History provides contextual details, plausibility tests, external consistency checks, and a role for free will. Though not proof of causation, correlation is a smoking gun; and history can often supply sufficient circumstantial evidence to convict.

The Problem of Causation

Economics is not the only place where correlation and causation get confused. Causality is a problem everywhere. For instance, physicians observe more heart attacks in people who are more obese and thus argue that obese people should diet to reduce the danger of heart failure. But is this really so? Perhaps people with weak hearts need more body fat, and dieting would worsen the danger of a heart attack. Or perhaps an unknown chronic viral infection causes both heart attacks and

[13] Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago, 1962).

[14] Fischer Black, "The Trouble with Econometric Models," *Financial Analysts Journal* 38, no. 2 (1982): 29–37.

[15] Jared Diamond and James A. Robinson, eds., *Natural Experiments of History* (Cambridge Mass., 2010).

body-fat accumulation, and dieting would only hide a cosmetic symptom of the virus while leaving it free to attack the heart.

Medical science infers causality with double-blind randomized trials. Equally, obese people must be randomly assigned to either a treatment group or a control group. People in the treatment group are put on a calorie-restricted diet and people in the control group are fed equally unpalatable food, designed to be indistinguishable from a diet.

First, assignment to groups must be utterly random. A caring physician might put patients she thought in dire danger of heart attacks in the treatment group, but that would spoil the test. If more dieters than control patients subsequently die of heart attacks, she cannot tell whether dieting killed them or prevented even more from dying. But if the assignment were utterly random, any difference in the death rates can be credited to (or blamed on) the treatment.

Second, neither the patients nor the physicians running the test may know who is in what group. People who know they have lost weight might act differently, or physicians might treat them differently, and either difference might cause a difference in outcomes between the two groups. Dieting must be the only difference between the two groups, otherwise some other unknown factor might be the true cause of any difference in outcomes between the two groups.

But if the test was done right, the difference between the treatment-group patients' old and new diets "caused" any difference in heart-attack rates between the treatment group and the control group. Such a difference-in-difference test allows a causal inference: putting obese patients on the diet prevents heart attacks.

Double-blind randomized trials are rare in economics. The divisions of Germany and Korea into capitalist and socialist halves might qualify as a test of socialism versus capitalism. The Iron Curtain arguably randomly assigned Germans to East and West Germany, and the Demilitarized Zone arguably did the same to Koreans. Prior to the late 1980s, neither set of leaders, nor even Paul Samuelson, could divine the victor in the cold war, so neither the citizens nor the economic policy makers on either side knew which was getting a treatment and which was getting a placebo.

Can we then conclude that different economic systems caused the differences in living standards evident by the late 1980s? Perhaps, but the Red Army seized northern Korea because the Japanese left an industrial infrastructure there, so the division was not truly random. The "treatment" might have been endogenous. Northern Koreans were more accustomed to factory work than their agrarian southern compatriots; but West Germany inherited a more comprehensive industrial base than did East Germany. East and West Germany differed in other ways.

For example, what became East Germany was mainly Protestant in 1945, while the future West Germany held a substantial Roman Catholic minority. Perhaps religious traditions, a latent factor, really caused any difference in economic prosperity.

Wherever genuinely randomized and double-blind trials occur, they are extremely useful. For example, Andrew Godley shows that Eastern European Jews who moved to London and New York at the turn of the century subsequently exhibited very different levels of entrepreneurship.[16] To the extent that the allocation of Jews to the two cities was random, this becomes a natural experiment on how environment differences affect entrepreneurship. Likewise, Peter Henry and Conrad Miller compare Barbados and Jamaica—Caribbean island nations with similar social, political, and economic institutions at independence, but with different development policies thereafter.[17] To the extent that their policy differences were random happenings, this was a natural experiment on how economic policies affect economic outcomes.

Unfortunately, such natural experiments are decidedly rare, so much causal inference in economics can be shaky. For example, economies with dynamic financial systems have been shown to grow faster.[18] Establishing this correlation was a useful exercise per se because it immediately falsifies any theories that imply a negative correlation or no correlation. But too many theories are consistent with a positive correlation, and so remain on the table. Does a dynamic financial system cause rapid growth? Or does rapid growth supercharge a country's financial system? Or does some other factor, a predominantly Protestant population, for example, cause both? This is more than an academic question, but multilateral financial institutions poured much money and effort into creating stock markets in post-socialist "transition" economies during the 1990s. Only if stock markets "cause" growth was this expenditure worthwhile.

Black's critique made economists and econometricians, in particular, keenly aware of the tenacious problems surrounding causal inferences. An arsenal of sophisticated techniques and penetrating insights has been deployed. However, impressive as they are, especially on a case-by-case basis, their limitations remain binding at a more general and collective level—as we argue below.

[16] Andrew Godley, *Jewish Immigrant Entrepreneurship in New York and London, 1880–1914: Enterprise and Culture* (New York, 2001).

[17] Peter Henry and Conrad Miller, "Institutions versus Policies: A Tale of Two Islands," *American Economic Review* 99, no. 2 (2009): 261–67.

[18] Robert King, and Ross Levine, "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics* 108, no. 3 (1993): 717–37.

Rummaging through the Toolshed of History

The great strength of the natural sciences is their basis in experiments conducted in controlled laboratory conditions. Randomized controlled experiments, usually on undergraduate subjects, can expose regularities in human behavior that usefully restrict the set of admissible theories; and the use of subjects in developing economies promises further insights but also raises new problems.[19]

But many of the deepest questions in economics concern whole nations and the dealings between them. The reader is invited to devise a controlled experiment to check whether or not bigger stock markets cause faster GDP growth. Electorates are disappointingly skeptical about letting economists use economies as laboratories to test unproven theories. And even when a theory is tested—say, Keynesian economics in the Great Depression or supply-side economics in the 1980s—we are rarely able to randomize or organize proper control samples. Economists can only look on with envy as a chemist fills two test tubes with the same reagent, treats one with a substance of interest, and notes the result.

The best economists can usually do in such circumstances is to find a useful *natural experiment*. Nature occasionally treats two otherwise identical groups differently in a way that resembles what economists would have done had they been allowed to run a controlled experiment. Such a natural experiment lets economists identify the causal effect of that treatment by measuring differences between the groups—first before and then after nature ran the experiment. The "difference in these differences" is plausibly caused by the different way nature treated the two groups.

Jared Diamond and James Robinson present several examples of such natural experiments that demonstrate the power of the technique. But they also warn that such cases are rare; and that apparent natural experiments can be invalidated by subtle initial differences between the groups, or by additional perturbations that affected them differently.[20]

For example, consider an economist searching for a natural experiment to ascertain the effect of a government policy with implications for the validity of an economic theory. Suppose the policy affects some people or firms more heavily than others. If the economist can sort the subjects in a way somewhat reminiscent of having randomly selected treatment and control groups, and then observe events unfold, causal inference is possible. The problem is finding a sorting mechanism that

[19] Colin Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, 2010); Angus Deaton, "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature* 48 (June 2010): 424–55.

[20] Diamond and Robinson, eds., *Natural Experiments of History*.

distinguishes heavily affected from lightly affected subjects in a way reminiscent of the randomization in medical trials. The groups must be identical in all other ways: the only permissible difference between them is that the policy weighs heavily on some and lightly on others.

The favored solution to this sorting problem is *instrumental variables*. This set of econometric techniques encompasses estimation using instrumental variables (IV) regressions, simultaneous equations (SE), generalized method of moments (GMM), and scores of related procedures. Though widely used, all these techniques are methodologically profoundly problematic. At least one valid instrumental variable must be found for each variable of interest in the estimation, and the criteria for validity are grueling. These are as follows:

1. Endogeneity and Exogeneity. A valid instrument must vary only in response to exogenous factors, that is, factors determined by nature, God, or people whose actions do not depend on the dependent variable in the model. In the medical trial, a random assignment of patients to the two groups serves as an exogenous way of distinguishing observations. An instrument also sorts observations by some criterion that is unaffected by the dependent variables the economist would test.

Economists often look to history here. For instance, countries' colonial histories and legal systems were shaped centuries ago, and so cannot be affected by their current economic performance. While instruments are sometimes taken from geography, linguistics, or other fields, economists seem happiest when rummaging about for instruments in history.

But does the arrow of time really make things so simple? James Tobin stresses that economics differs fundamentally from the natural sciences because people's economic decisions depend on their expectations of future events; while the actions of pendulums, atoms, and planets do not.[21] This teleological quality at the very heart of economic theory means that the future "causes" the present in economics. For example, shareholders' expectations about future dividends determine a stock's price today. Can such temporal ricochets affect the flow of history in general?

Let us explore colonial origin. If British, French, Spanish, and Portuguese colonies were scattered randomly throughout the world, colonial heritage would qualify as an exogenous instrument. But France lost Canada in 1759 and abandoned the colony in 1763, demanding instead the sugar island Guadeloupe as the price of a peace treaty with Britain. British government officials disproportionately chose to make

[21] James Tobin, "Money and Income: Post Hoc Ergo Propter Hoc?" *Quarterly Journal of Economics* 84, no. 2 (1970): 301–17.

and defend claims of sovereignty over territories with agricultural potential; France, Spain, and Portugal for the most part did not.

Is a British colonial heritage then the "cause" of Canada's agricultural exports? Is a French colonial heritage the cause of Guadeloupe's economic dependency? Perhaps; but Britain and France deliberately colonized places with certain characteristics, like physicians choosing patients with certain characteristics for their trials and thereby invalidating the initial randomization. How do we know that Canada's agricultural potential didn't cause it to end up under British suzerainty? Such questions may be answerable, but their asking demonstrates that historical variables, even very deep ones, are not a priori exogenous.

Careful researchers must thus work hard to validate their exogeneity assumptions. One approach is a careful reading of the historical record surrounding the data used to construct the instrumental variable.[22] A nonrandom initial difference between subjects might become evident over time; and another perturbation might affect different subjects differently. Either could confound the natural experiment into presenting a false picture of what causes what.

2. Weakness and Strength. A valid instrument must be strongly correlated with the treatment. Economists generally cannot randomly assign observations to treatment and control groups; the instrument must do this. For example, an economist might be interested in how comparable-worth wage laws affect unemployment, but is worried that unemployment might also affect a country's labor laws. The economist therefore rummages about in history for an instrument and, let us suppose, selects the longitude of each country's capital city.

This variable might meet the endogeneity criterion described above, but it is no good as an instrument unless it correlates strongly with the treatment. After all, its purpose is to randomly allocate countries to the treatment group, those with comparable-worth laws, and the control group, those without such laws. Longitude can hardly do this if it is uncorrelated with the presence of those laws.

James Stock and Mark Watson ascertain that instrumental variables achieving a joint F statistic below ten in a regression explaining the relevant treatment variable may have a *weak instruments problem*.[23] Though they provide techniques for using weak instruments nonetheless in certain situations, failure to pass a weak instruments test generally consigns otherwise commendably instrumental variables to the dustbin of econometrics.

[22] Abhijit Banerjee and Lakshmi Iver, "Colonial Land Tenure, Electoral Competition and Public Goods in India," in *Natural Experiments of History*, ed. Jared Diamond and James A. Robinson (Cambridge, Mass., 2010), ch. 6.

[23] James Stock and Mark Watson, *Introduction to Econometrics*, 2nd ed. (Boston, 2007).

Dismayed at longitude's failing this test, the persevering economist might rummage further and, after a hundred or so tries, find the cosine of mean-squared 1880s rainfall correlating with a dummy for comparable-worth laws ($p < 1\%$). Unfortunately, a variable, even a serenely exogenous one, that correlates with the treatment only incidentally, and after days of rummaging through the toolshed, is really merely a selected reflection of the treatment variable itself. Any endogeneity problems that afflict the original variable afflict its reflection too.

Searching for false positives is no way to uncover strong instruments. We do not charge economists with rifling through history for Type II errors, but worry that editors and referees tempt authors by demanding that they force causally circular data into inappropriate square instrumental variables econometrics.[24]

Weak instrument problems are especially likely to arise if the data are noisy—that is, observed imperfectly. For example, a highly acclaimed and carefully done study by Daron Acemoglu, Simon Johnson, and James Robinson uses mortality rates of early colonial settlers as an instrumental variable to sort countries by propensity to establish property rights protecting institutions.[25] If settlers were initially randomly distributed across colonies, and property rights protecting institutions were in greater demand where more settlers survive, this variable qualifies as exogenous. Nonetheless, a well-articulated debate between David Albouy and Acemoglu, Johnson, and Robinson about the accuracy of historical mortality rates demonstrates how data uncertainties can create a weak instruments problem even if the instrument is plausibly exogenous.[26]

3. Latency and Blatancy. A valid instrument must not be thrown off by latent factors. The increasing popularity of historical variables as instruments makes this a growing problem.

There are many important cases where colonial origin, legal-system origin, religious history, settler mortality, and the like are arguably exogenous and are correlated with treatment variables of interest. For example, accepting that the origin of a country's legal system cannot be

[24] David Weimer, "Collective Delusion in the Social Sciences: Publishing Incentives for Empirical Abuse," *Review of Policy Research* 5, no. 4 (1986): 705–8.

[25] Daron Acemoglu, Simon Johnson, and James Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91, no. 5 (2001): 1369–401.

[26] David Albouy, "The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data," National Bureau of Economic Research working paper no. 14130, 2008; Daron Acemoglu, Simon Johnson, and James Robinson, "A Response to Albouy's 'A Reexamination Based on Improved Settler Mortality Data,'" mimeo (Mar. 2005); Daron Acemoglu, Simon Johnson, and James Robinson, "Reply to the Revised (May 2006) version of David Albouy's 'The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data'" (2006).

caused by its current financial dynamism, suppose an economist finds significantly more dynamic financial systems in common law countries. She rightly uses legal origin as an instrument for financial development; that is, she uses legal origin as an exogenous criterion for sorting countries in a way that also likely ends up sorting them by financial development. Then she can test whether the common law countries, which have more dynamic financial sectors purely by dint of having common law legal systems, grow faster than otherwise identical countries that lack dynamic financial systems purely by dint of lacking common-law legal systems. If she includes appropriate control variables, so the countries truly are otherwise identical, this is arguably a valid test, and she can conclude with a straight face that financial development causes growth.

Now suppose another economist wants to see if agricultural productivity causes economy growth, and finds the latter variable also correlating highly with legal origin. The second economist, using legal origin as an instrument, regresses economic growth on agricultural productivity; and, finding a significant coefficient, concludes that agricultural productivity causes growth.

This, unfortunately, does not fly. The second economist should have read the literature—in particular, the first economist's paper. He knows financial development matters in this setting, and has a latent variable problem in his regressions, unless he includes that variable too. Moreover, publication of the second economist's paper means the first economist's article is no longer convincing as regards causality. She now has a latent-factor problem, for she failed to control for agricultural productivity, an endogenous variable that the second economist proved to be important. The key point here is that each subsequent paper that reuses an instrument in a shared context contributes an additional latent factor problem to all the existing studies.

Tragically, commonly used instrumental variables lose value with overuse. This is because the instrumental variables are *nonexclusionary* (the first economist to use an instrument cannot prevent others from using it too) and can be *rivalrous* (each successive use potentially compromises the instrument's validity in every previous and subsequent use). Absent a comprehensive multinational agreement enforcing their patenting, instrumental variables are stymied by a classic Tragedy of the Commons.[27]

4. An Econometric Tragedy of the Commons. The requirements of exogeneity, strength (no weak-instruments problems), and blatancy (no latent-factor problems) severely limit the supply of valid instrumental variables. This leads to their recycling. Each individual study

[27] Garrett Hardin, "The Tragedy of the Commons," *Science* 162, no. 3859 (1968): 1243–48.

may look econometrically rigorous—its instruments exogenous and strong. But authors of literature reviews, who must evaluate the collective contributions of many such studies, cannot but doubt the validity of each study, given the others.

Economists have long stressed internal consistency. An economist generally may not begin a proof assuming a logarithmic utility function and then switch to a constant elasticity of substitution (CES) utility function partway through.[28] But even the best economics journals have no problem with logarithmic utility in one article and CES in the next, even if each article is utterly devastated by the assumption used in the other.

This lack of concern for external consistency is a challenge to theorists, but a disaster for empirical economics when issues of causation arise. An effect that is blatantly significant in one study is necessarily potentially latently significant in all others that explore the same economic questions, and probably in studies that examine many related economic questions too. Individual articles can sustain a veneer of consistency, but the collective literature cannot.

A Tragedy of the Commons has led to an overuse of instrumental variables and a depletion of the actual stock of valid instruments for all econometricians. Each time an instrumental variable is shown to work in one study, that result automatically generates a latent variable problem in every other study that has used, or will use, the same instrumental variable, or another correlated with it, in a similar context. We see no solution to this. Useful instrumental variables are, we fear, going the way of the Atlantic cod.

Learning from Repeating History

Fortunately, there are ways we can learn about causation from history without rummaging for instrumental variables. A prime example of this is the event studies of financial economics. A second is Granger causality (*G*-causality) tests, widely used by macroeconomists.

Event Studies. Event studies are perhaps the most direct test for causality available to economists.[29] For example, a financial economist who wanted to see if comparable-worth laws add value to firms might identify the precise dates on which each U.S. state with such laws first

[28] Logarithmic utility assumes a subject's utility (hedonic pleasure) from consuming C to be a function of the form $U = a \log(C)$, while constant elasticity of substitution utility assumes $U = C^{1-a}/(1 - a)$. The two are equivalent if a is 1, but not otherwise. Economic theorists often choose a functional form to make the algebra easier; however, results based on one form often do not follow if another is used instead.

[29] John Campbell, Andrew Lo, and A. Craig MacKinlay, *The Econometrics of Financial Markets* (Princeton, 1997), ch. 4.

https://doi.org/10.1017/S000768051100002X Published online by Cambridge University Press

announced them. If the value of a portfolio containing the stocks of all the firms operating in the announcing state rises significantly relative to the value of a portfolio containing all other stocks on each such event date, the financial economist is on passably solid ground inferring that comparable worth "causes" increased firm values.

The power of event studies lies in repetition of history. If each of a large collection of economically similar events corresponds to similar patterns in the data, we can infer that something significant is happening. In this example, each state's announcement repeats the event, and if each repetition is associated with a similar relative stock value hike for the firms in the affected state, a pattern is evident and causality can be inferred.

An inference of causality is justified by Occam's razor: that the legal reform causes stock prices to change is reasonable because the reverse is manifestly implausible. For stock price hikes to cause the laws, state legislators would have to patiently monitor the ticker tape until a day when the stocks of firms in their state, and only those stocks, rise; and then burst forth with news of new labor laws.

However, even here, we must beware of latent factors. For example, if states tend to adjust their minimum wages whenever they adopt comparable-worth laws, the minimum wage might be causing the stock-price changes.

Also, insignificance in an event study cannot prove an absence of causation, for economic decision-makers' expectations of the future again come into play. If Iowa's adoption of comparable-worth labor laws were all but assured months ahead of their actual unveiling, the unveiling would not move stock prices. Investors would long ago have adjusted their expectations about the dividends of Iowa firms, and little or nothing would happen when those expectations were realized.

The event study technique is thus weakened by investors' collective learning. But learning is usually incomplete—as long as some probability of history following an alternative path remains nonzero until the event actually occurs, event study can be informative about causality. Moreover, many interesting economic phenomena are fundamentally amenable to perfect prediction by neither econometricians nor the people they model.[30] The unfolding of history reveals new information, and

[30] C. R. Nelson, "The Prediction Performance of the F.R.B.-M.I.T.-Penn. Model of the U.S. Economy," *American Economic Review* 62 (1972): 902–17; Richard Roll, "R-Squared," *Journal of Finance* 43, no. 3 (1988): 541–66; Francis Diebold, "The Past, Present and Future of Macroeconomic Forecasting," *Journal of Economic Perspectives* 12, no. 2 (1998): 175–92; Ricardo Caballero, "Macroeconomics after the Crisis: Time to Deal with the Pretence-of-Knowledge Syndrome," MIT Department of Economics Working Paper No. 10-16, 2010; and others.

human ingenuity creates innovations—neither, by definition, is predictable; yet both are central to economics.

Granger Causality Tests. Something akin to an event study is sometimes econometrically feasible in panel data. Granger causality tests exploit a definition of causal relations between random variables proposed by Norbert Wiener: one variable "Granger-causes" (or "*G*-causes") another if a forecast of the second variable based only on its past values is made significantly more accurate by using past values of the first variable as well.[31]

In practice, these forecasts are almost always linear regressions, so the test is really about one variable "*G*-causing" another if a regression of the latter variable on its own past values and past values of the former variable has a significantly higher R^2 than a regression of the latter variable on its own past values alone.

For the test to be valid, both variables must be stationary—they must not have a common trend. Trends are removed by taking first differences, second differences, or if necessary, even higher-order differences, until a panel of stationary data is obtained. This is reasonable, for if one variable causes another, changes in the first variable presumably also cause changes in the second.

Like other tests of causality, this approach requires that the economist worry about latent factors, for if a third variable "causes" both variables being tested for Granger causality, a false positive can result. And, as in the case of event studies, an absence of evidence of causality is not evidence of its absence.

Granger causality tests are perhaps uniquely vulnerable to the fundamental teleology of economic theory. If central bankers adjust the money supply based on their expectations of future GDP growth, a Granger causality test might erroneously show the money supply "causing" GDP growth. Because economics is about people's decision-making under uncertainty, expectations about the future cause present decisions. If those expectations turn out to be correct in general, the future can seem to cause the past.[32]

Event studies are less vulnerable to this critique because stock prices can be observed at very high (daily and intraday) frequency and, if announcement times are sufficiently precise, Occam's razor can cut away alternative causality scenarios. For example, firms usually announce major strategic decisions after the stock exchange closes for the

[31] Clive Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica* 37 (1969): 424–38; Norbert Wiener, "The Theory of Prediction," in *Modern Mathematics for Engineers I*, ed. E. F. Beckenbach (New York, 1956).

[32] John Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica* 29 (1961): 315–35.

day. An event study of firms' announcements of diversifying takeovers finding their stock price the next day significantly below the closing price just prior to these announcements is consistent with diversification causing shareholders to revise downward their estimates of the firm's value. Reverse causality would entail CEOs, foreseeing stock-price drops, deciding to announce diversifying takeovers. This is not impossible, but it is implausible. Economic theory provides many reasons for diversification to destroy value, but no reasons for CEOs to act as reverse causality would demand.

However, Granger causality can work where event studies do not. Event studies can be impractical if the variable of interest is observed only at a low frequency (quarterly or annually) and a long enough time series to permit meaningful statistical tests does not exist. Moreover, if the variables of interest exhibit sluggish adjustments or are obscured by substantial noise, as many macroeconomic variables and product prices can be, Granger causality tests can fail to detect bona fide causal relations.

Implausibly Deniable Causality

Absence of evidence of a given direction of causation is not evidence of its absence, and is certainly not evidence of causation in the reverse direction. Neither instrumental variables regressions, nor event studies, nor Granger causality tests can assert an absence of causal connection. That a negative cannot be proven is an epistemological truism, but that doesn't prevent economists from trying.[33]

Statistical insignificance in an event study does not mean the events definitively do not cause changes in stock prices. The event dates might be insufficiently precise, or stock prices might be too volatile to detect the signal reliably, or investors might have expected the event with sufficient probability that its price impact was negligible. Granger causality tests can also be muddied by the timing of expectations revisions, by noisy data, and by insufficiently long or excessively persistent panel data.

An absence of significance in an instrumental variables framework likewise does not mean an absence of causality. The instrument may not be strong enough, latent variables may lie hidden in the statistical background, or the effect may be obscured by the noise. Even more important, an absence of significance in an instrumental variables framework does not imply reverse causality. Proving reverse causality requires

[33] Lawrence Summers and Robert Stambaugh, "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance* 41, no. 3 (1986): 591–603.

specifying a regression that represents the reverse causality, complete with its own control variables and exogenous strong instruments for its endogenous right-hand side variables.

Dusting Off History

History ought to be intrinsically interesting to economists. Economics seeks to explain patterns in the progress of individuals and collectives—communities, corporations, and nations. History documents the past that generated economists' datasets, and so ought to arouse economists' intellectual curiosity. But we propose that the study of history offers economics much more.

History provides context—an intensity of information around a few observations—and this can sometimes be as useful as a large dataset. A good example of this is Alfred Chandler's *Strategy and Structure: Chapters in the History of American Industrial Enterprise* (1962). This work lays out, in intricate detail, the inner workings of DuPont, General Motors, Sears, and Standard Oil as they adopted a new corporate structure that he dubs the M-form. The degree of detail, based on careful documentation of how key decisions came to be made, shows that the corporations' strategies must determine their structures, not the converse. These observations continue to shape studies of business strategy, and much recent work also applies Chandler's strategy for ascertaining patterns of causality. For example, Geoffrey Jones and Tarun Khanna, surveying the business history literature, point out how historical information on early European multinationals illuminates underlying causes of their diversification and development into business groups.[34]

Historical studies have a collective methodology: external consistency matters. History subjects competing narratives to ongoing tests of plausibility, and this narrative format forces an external consistency. To sustain credibility, a good historical narrative must connect the "dots" of all relevant historical events with causal links. And while historians debate the importance of individuals as opposed to impersonal forces, history is more amenable to the concept of free will than is neoclassical economics; and causality is far more interesting if there is free will. In sum, we believe more attention to history offers economists more defensible arguments about causality.

The Importance of Context. Economics strives for simplification that reveals underlying causal principles. The detail and contextualization favored by historians complicates economists' models. While some

[34] Geoffrey Jones and Tarun Khanna, "Bringing History (Back) into International Business," *Journal of International Business Studies* 37 (2006): 453–68.

historians can be accused of excessively imaginative reconstruction of causality and deliberately biased searches for historical evidence supporting their favored narratives, economists are hardly immune to mistaken musings and confirmation bias. But historians' purpose is, first and foremost, a sustained effort to reveal causality. That shared purpose makes history intrinsically interesting to economists.

Historical studies about economic and financial events offer chronological sagas of unfolding developments. They link outcomes to events, reactions to actions, and (perhaps most crucially to economists) historically consequential errors to critical decision-makers' private preferences and incomplete information. History is composed of narratives that "connect the dots" in causal terms.

History, unlike economics, pays great attention to external consistency. Historians' narratives gain credibility by their finesse at connecting all the dots. This attention to context can be illuminating.

For example, Germany and Japan are "bank-based" economies: their big businesses rely on banks for capital and seldom issue new shares onto their stock markets. In contrast, Anglo-Saxon countries are "stock-market-based" economies: their big companies rely extensively on share issues to finance growth, and long-term bank loans are markedly less important. An econometrician would correctly detect no indication that one system causes higher living standards than the other. However, a historian might dissent. Both Japan and Germany industrialized in the late nineteenth and early twentieth centuries, and both were stock-market-based economies in their high-growth decades.[35] Banks rose to dominance amid Japan's postwar reconstruction and under Germany's National Socialist government, though Bismarck began shifting German regulations toward favoring banking much earlier.[36] Indeed, that any major economy has ever industrialized successfully without a large stock market is unclear.[37]

This example highlights the importance of path dependence. Germany and Japan both had to finance costly large-scale postwar reconstruction, and both used vastly expanded banking systems to do so. Path dependence tends to undermine assumptions of ergodicity, the premise

[35] Caroline Fohlin, "The History of Corporate Ownership and Control in Germany," in *A History of Corporate Governance around the World: Family Business Groups to Professional Managers*, ed. Randall K. Morck (Chicago, 2005), 223–77; Randall K. Morck and Masao Nakamura, "Business Groups and the Big Push: Meiji Japan's Mass Privatization and Subsequent Growth," *Enterprise and Society* 8, no. 3 (2007): 543–601.

[36] Ranald Michie, *The Global Securities Market: A History* (Oxford, 2008).

[37] See Raghuram Rajan and Luigi Zingales, "The Great Reversals: The Politics of Financial Development in the Twentieth Century," *Journal of Financial Economics* 69, no. 1 (2003): 5–50. Even communist China has established stock markets. Their contribution toward that country's further development remains to be seen.

that time-series and cross-section variations are statistical substitutes. In this case, the cross-section is silent, but a few historical observations are informative.

By putting their current financial systems in context, history gives economists a better understanding of their data. The detailed economic histories of Japan and Germany are case studies, not data. But their wealth of detail provides a context in which to evaluate broader hypotheses and disentangle the effects of path dependence. For example, Stephen Haber's recent comparative description of the development of banking in the U.S., Brazil, and Mexico does precisely this.[38] The value of descriptive history in addressing these sorts of issues is surveyed by Jones and Khanna, and reiterated by Morck and Yeung.[39] In this way, a few observations—perhaps even just one—can provide an intensity of information that allows inferences even a large dataset might not reveal.

Competing Narratives and Occam's Razor. Such exercises are useful to economics because the uncovering of previously unknown historical evidence and the unfolding of current events into the tapestry of history provide ongoing tests of competing narratives. Occam's razor shapes the tapestry: narratives rendered less plausible fall away before narratives rendered more plausible.

History thus has its own way of ascertaining validity. A historical narrative must be logical and backed by evidence. Historians construct, modify, extend, and prune their narratives to maintain internal and external consistency. Sometimes this reinforces established narratives; at other times it leads to their replacement by another narrative in a process, much as new paradigms overturn old ones in the sciences.[40] In both cases, old paradigms can be tenacious, and perhaps hang on longer than they should. Indeed, really major changes must often await a new generation of scholars with less human capital invested in the old paradigm. Thus Samuelson's famous quip: "funeral by funeral, economics does make progress."[41] This happens in the sciences, too: quantum mechanics took over physics, not because many physicists changed their minds, but because old physicists retired and young physicists found the new paradigm convincing.[42] Evolution took even longer to become

[38] Stephen H. Haber, "Politics Banking, and Economic Development: Evidence from New World Economies," in *Natural Experiments of History*, ed. Jared Diamond and James A. Robinson (Cambridge Mass., 2010), ch. 3.

[39] Jones and Khanna, "Bringing History (Back) into International Business," 453–68; Randall K. Morck and Bernard Yeung, "History in Perspective: Comment on Jones and Khanna, 'Bringing History (Back) into International Business,'" *Journal of International Business Studies* 38 (2007): 357–60.

[40] Kuhn, *The Structure of Scientific Revolutions.*

[41] Quoted in Wilson, *Consilience: the Unity of Knowledge*, 52.

[42] Helge Kragh, *Quantum Generations: A History of Physics in the Twentieth Century* (Princeton, 2002).

the central paradigm of biology.[43] Economic theories of monopoly, macroeconomics, and individual choice, to name but a few, have undergone similar transformations, and some of these may well have required funerals, or at least retirements, to take hold. History can sometimes help the upstarts, when business historians show U.S. students of multinationals that European companies in the nineteenth century were as enthusiastic multinational investors as their U.S. counterparts in the twentieth century.[44] Similarly, Chandler's pioneering work on the importance of economies of scale and scope dominated the field for a generation, but the data ultimately led Philip Scranton to showcase the persistent importance of specialized production, alongside mass production, in propelling U.S. industrialization in the late nineteenth and early twentieth centuries.[45] Chandler's finding that U.S., U.K., German, and Japanese firms progressed from family control to the stewardship of professional managers likewise caused a generation of economists to view this sequence as the baseline paradigm of business everywhere. This too was qualified by historical work showing those four countries to be atypical, and demonstrating that ongoing family control over large business empires continues to be the norm in most countries.[46] Yet another example is how Henry Ford's philosophy of management remained broadly influential until business historians entered the debate.[47]

The credibility of each narrative depends not only on its ability to "connect the dots" between past events, but also to explain new dots that arise from archaeological digs, previously forgotten archives, and the unfolding of history from current events. These tests are not econometric, but they are powerful nonetheless. Narratives that were once deeply compelling can be cast aside when they fail to connect important dots. For example, the narrative of Western colonialism civilizing the benighted savages of Africa and Asia could not connect the dots of two world wars, and is now itself an historical curiosity.

The connecting of such dots can be every bit as painstaking as the careful assembly of a large econometric database. For example, Stanley Engerman and Robert Fogel assembled historical data on slaves in the

[43] Edward Larson, *Evolution: The Remarkable History of a Scientific Theory* (New York, 2004).

[44] Wilkins, *The Emergence of Multinational Enterprise*; Hertner and Jones, *Multinationals*; Jones, *Multinationals and Global Capitalism*; and others.

[45] Philip Scranton, *Endless Novelty: Specialty Production and American Industrialization, 1865–1925* (Princeton, 1997).

[46] Randall Morck, ed., *A History of Corporate Governance around the World: Family Business Groups to Professional Managers* (Chicago, 2005).

[47] Steven Tolliday and Jonathan Zeitlin, eds., *The Automobile Industry and Its Workers: Between Fordism and Flexibility—Comparative Analysis of Developments in Europe, Asia, and the United States from the Late Nineteenth Century to the Mid-1980s* (New York, 1987).

American South, and argued that their owners took good care of their property to maintain its value, as economic theory would predict.[48] A spirited dispute followed over the quality of their historical data.[49]

A powerful example of historians connecting causal dots is Charles Kindleberger's historical analysis of financial manias, panics, and crashes.[50] Kindleberger sets out detailed histories of each major financial crisis from the advent of modern stock markets in the early 1600s to the 1970s. He distills from these histories a common trajectory that each crisis follows: an economic dislocation that creates genuine economic profit opportunities, an inrush of capital to fund them, a popular demand for deregulation to allow broader participation, a continued capital inflow after the profit opportunities are exhausted, manic episodes of capital chasing illusory high returns from stock markets to commodities to real estate, a crash, and a popular fury with financiers that usually heralds tough new regulations—which persist until the next cycle. The neat obedience of all subsequent financial crises to Kindleberger's thesis enhances its credibility. Alternative narratives based on stock-market efficiency have fallen aside, and Kindleberger's remains the "narrative to beat."

A Broad-Minded Consistency. History is a correspondence between individuals, generations, and eras, in which one writer cannot easily ignore the scrawls of the others. The last point in particular contrasts starkly with economists' precise attention to the internal consistency of every article, rather than external consistency between studies. Above, we stressed that using a variable on both the left-hand and right-hand sides of OLS regressions seriously bothers economists if done within an article; but not if a few pages of references, a title, and an abstract intervene. This narrow-minded consistency is more than an econometric problem.

Our reading of the literature suggests that historians can be more broad-minded about consistency. More respect for history would, we think, promote a long overdue regard for external consistency across studies in economics. Good historians connect the dots across broad patterns of human endeavor. Even historians focused on a relatively narrow national or temporal band must connect facts in geography to facts in politics, climate history, psychology, and (of course) economics. This expanse of context is rare in economics.

[48] Stanley Engerman and Robert Fogel, *Time on the Cross: The Economics of American Negro Slavery* (Boston, 1974).

[49] Robert Fogel, *The Slavery Debates, 1952–1990: A Retrospective* (Baton Rouge, 2003); Herbert Gutman, *Slavery and the Numbers Game: A Critique of* Time on the Cross (Champaign, Ill., 2003).

[50] Charles Kindleberger, *Manias, Panics and Crashes* (New York, 1978).

For example, development economics was long founded on the premise that poor countries were basically like the United States, but poorer.[51] This perspective justified massive foreign aid. When this effort succumbed to widespread corruption, attention turned to structural reforms designed to make developing countries more like poor versions of the United States, so that future aid initiatives might find better traction. This drastically oversimplifies a complicated field of economics, but we believe the simplification captures something essential: a lack of concern for external consistency.

Historians studying the problem of persistent poverty provide more context, and this lets them expose interesting patterns that can be checked for consistency across many similar historical events. For example, Haber, writing on Latin America, chronicles episodes of aborted industrialization, and discerns a pattern: the region's elites are enriched by industrialization, but fear losing control should institutions ever develop fully.[52] Haber, Douglass North, and Barry Weingast, and North, John Wallis, and Weingast draw from the histories of many countries to document patterns that consistently distinguish developmental success stories from developmental failures.[53]

While such economic historians rely on econometric evidence where it is credible, their narratives do not rely fundamentally on *F*-tests or likelihood ratios. Their claim to legitimacy is that they start from detailed information-rich case studies, connect the dots to discern plausible patterns of causality, and demonstrate a generality to these patterns by demonstrating a broader external consistency with collected previous works.

Taking Free Will Seriously. Economics was deeply affected by the philosophy of causal determinism, which the natural sciences embraced throughout the nineteenth century. That philosophy is most famously espoused by the philosopher Pierre-Simon Laplace thus:

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing

[51] Deepak Lal, *The Poverty of "Development Economics"* (Cambridge, Mass., 1985).

[52] Stephen H. Haber, *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800–1914* (Stanford, 1997).

[53] Stephen Haber, Douglass North, and Barry Weingast, *Political Institutions and Financial Development* (Stanford, 2008); Douglass North, John Wallis, and Barry Weingast, *Violence and Social Orders* (Cambridge, U.K., 2009).

> would be uncertain and the future just like the past would be present before its eyes.[54]

For this intellect, dubbed "Laplace's demon," every event is a cog in a mechanical chain stretching back to the beginning of the universe.

The neoclassical synthesis of the 1870s, which still largely defines microeconomic theory, drew heavily from the physics of the time and presents human beings as part of this cosmos.[55] Human beings are causally deterministic utility-maximizing machines, whose decisions are fully determined by their predefined preferences and budget constraints, which are fully determined by a mechanical chain stretching back into the depths of time. In such a world, causation is both simple and uninteresting, for nothing is exogenous except the prime mover who set the clockwork moving eons ago. Yet, this analytical framework came to guide causal interpretations of inputs, changes, and outputs in the econometrics of the Age of Data.

In truth, economists have never really accepted causal determinism. Even the most committed neoclassicists contemplate exogenous interventions: Acts of God, and even policy changes, that somehow originate outside such rows of dominos, and that send deterministic rows of utility-maximizing human decisions toppling down alternative paths.

Physics long ago abandoned causal determinism; indeed, quantum mechanics left it no choice by adding intrinsic uncertainty to time and space. This, in turn, freed philosophy to contemplate human free will. Economics hardly noticed these changes. Yet if free will exists, human decisions must be exogenous in the deepest philosophical meaning of the term, and the origins of all economically interesting causal chains of events.

Historians have long argued about the importance of individuals, as opposed to deterministic forces. If free will matters, individuals are important. The cognitive processes, emotions, compulsions, and desires within human decision-makers are the ultimate causes of the phenomena economists study.

History records autobiographical and biographical information that can tell us what people were thinking, worrying about, or pursuing when they did what they did. Perhaps economists might investigate these records to see what they reveal about what caused key decision makers to decide as they did. Fundamental advances in understanding

[54] Pierre-Simon Laplace, *Essai philosophique sur les probabilités* (1814), transl. by Frederick Truscott and Frederick Emory, *A Philosophical Essay on Probabilities* (Dover, U.K., 1951).

[55] Philip Mirowski, *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics* (Cambridge, U.K., 1991).

phenomena like entrepreneurship can emerge from ascertaining the constraints, knowledge, motives, and cognitive processes of those key decision-makers.[56]

Cognitive dissonance and other behavioral biases surely cause people to misremember such things ex post, and even to lie about them deliberately. But the historical record contains real-time archives that can occasionally reveal the sometimes uncomplimentary motives that caused particular chains of events to unfold. Of course, archives can be biased, deliberately manipulated, or released selectively, and careful business historians are alert for this; but archives can also upend aged decision-makers' sanitized accounts.[57]

Conclusion

We conclude that Black's critique of econometrics, his entirely reasonable argument that correlation is not causation, may well have been taken too seriously by economists. As Edward Tufte equally reasonably points out, "Correlation is not causation but it sure is a hint."[58] More precisely, correlation is a necessary, but not sufficient, condition for causation. This makes tests for correlations in economic data important. Econometric tests for causality may well be much less useful, for they can often be extraordinarily difficult to do well. The progress of economics may well be better served by careful and reliable tests for correlations than by flawed tests asserting or denying causality. How then can economists ascertain what causes what? Here we conclude that economists might make better use of history. History is far more than a toolshed for instrumental variables. History is filled out with nuances that contextualize events. History is composed of competing narratives that must "connect the dots" or lose credibility. History records autobiographical and biographical information that can tell us what people were thinking, worrying about, or pursuing when they did what they did. History is a correspondence between individuals, generations, and eras, in which one writer cannot easily ignore the scrawls of the others.

Popper and especially Lakatos argue that science progresses by the successive falsification of whole theories, not individual hypotheses.[59]

[56] Mark Casson, Bernard Yeung, Anuradha Basu, and Nigel Wadeson, eds., *Oxford Handbook of Entrepreneurship* (Oxford, 2006); Mark Casson, "Entrepreneurship," in *The* Fortune *Encyclopedia of Economics*, ed. D. R. Henderson (New York, 1993).

[57] Richard Cox and David Wallace, eds., *Archives and the Public Good: Accountability and Records in Modern Society* (Westport, Conn., 2002).

[58] Edward Tufte, *The Cognitive Style of PowerPoint* (Cheshire, Conn., 2003).

[59] Imre Lakatos, *Proofs and Refutations* (Cambridge, U.K., 1976).

This is why a broader respect for external consistency is needed if economics is ever to gain acceptance as a science. This is also why economics must come to grips with the fact that its observations are usually context dependent. Statistical tests for causality are obviously useful once a theory has been enunciated, but contextualized observation is more often the source of the broad pictures and frameworks that coalesce into the theories we test—in science and economics.[60] Indeed, Adam Smith built his theories, arguably the basis of the whole of modern economics, around detailed, qualitative observations of the workings of a pin factory.[61]

Econometrics has served economists well, and it continues to do so. But it cannot answer every question, and has especially intractable problems with many questions of causation. We do not call for any unwinding of past work, but for a reinvestment in history, so that the complementary relation between statistical analysis and historical investigation we describe above can step in where econometrics falters.

A natural complementarity portends benefits both economists and historians; but we (as rational and self-interested economists) perceive primarily the benefits to our field. Economics as a discipline has standardized a powerful methodology, which may indeed be useful in other fields.[62] Relying on theories of constrained optimization and equilibrium, tempered by behavioral regularities and the availability of information, economics builds empirically falsifiable statements and guides the collection and interpretation of historical information. Some of these statements are readily amenable to econometric tests, but others—especially those about one thing causing another—are more difficult to test. We argue that economists can in turn look to history for help here. Economists already make use of repetitions of history in the forms of event studies and Granger causality tests. But economists might also gain insights about causality by attending to details of context, weighing the plausibility of competing narratives, assessing external consistency, and studying the constraints, motives, and recollections of key decision-makers—either directly or through archives. All these methodologies surely also have their problems too. But we believe them to be less critical than the difficulties inherent in using instrumental variables methods to assess causation in many important settings.

[60] Paul Feyerabend, *Against Method: Outline of an Anarchistic Theory of Knowledge* (London, 1975); Jones and Khanna, "Bringing History (Back) into International Business," 453–68.

[61] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (London, 1776).

[62] Lazear, "Economic Imperialism," 99–146; Lamoreaux, Raff, and Temin, "Economic Theory and Business History," in *The Oxford Handbook of Business History*, ed. Jones and Zeitlin, ch. 3.

. . .

RANDALL MORCK is Distinguished University Professor and Jarislowsky Chair in Financial Economics at the University of Alberta, Canada, and research associate at the National Bureau of Economic Research. His publications include *The History of Corporate Governance around the World: Family Business Groups to Professional Managers*, which he edited in 2005.

BERNARD YEUNG is dean and Stephen Riady Distinguished Professor in Finance and Strategic Management at the National University of Singapore Business School. He has written extensively about corporate governance, foreign direct investment, and business history, including most recently (with Randall Morck and Deniz Yavuz) "Bank Ownership, Capital Allocation, and Economic Performance" in the *Journal of Financial Economics* (2011).