EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 17(1)
March 2020: 228–241

Edward Leamer Deserves a Nobel Prize for Improving Argumentation That Uses Statistics

Arnold Kling[1]

LINK TO ABSTRACT

This book could be classified under the heading of "metastatistics." Statistics is the theory of inferences ideally *drawn from data. Metastatistics is the theory of inferences* actually *drawn from data… Metastatistics analyzes how the researcher's motives and opinions influence his choice of model and his choice of data… it also deals with the social mechanism by which information is transmitted among individuals.*

—Edward Leamer (1978, v)

Edward Leamer deserves the Nobel Prize in Economic Sciences for launching the movement to examine critically the uses of statistical methods in empirical research. The movement has had repercussions that go beyond econometrics. It has affected medicine and epidemiology, where John P. A. Ioannidis has been a leading figure in pointing out methodological failures (Ioannidis 2005; 2016; Begley and Ioannidis 2015). It has impelled psychology and behavioral economics to confront what has become known as the 'replication crisis' (Camerer et al. 2018).

Leamer noted that economists usually work with data that is observational, not experimental. With observational data especially the proper specification of the analysis is uncertain. To deal with mis-specification risk, researchers employ what Leamer (1978) termed "specification searches." Leamer rigorously analyzed

1. Mercatus Center at George Mason University, Arlington, VA 22201.

the pitfalls of this process, eventually leading economists to rethink their approach to doing empirical work.

Suppose that you wish to compare the performance of charter schools with public schools by looking at test scores for students in each type of school. How do you control for the way that test scores are affected by the characteristics of the students enrolled in the schools, not just the differences in the approaches to education taken by the two types of schools?

Before the publication of his 1978 book *Specification Searches* and his 1983 article "Let's Take the Con Out of Econometrics" in the *American Economic Review*—two works which as of February 2020 together had nearly 6,000 Google Scholar citations (**link**)—the standard method was multiple regression. The investigator would have estimated a regression equation to explain student test scores using the school type (charter or public) while also including average income of parents, average education of parents, and other auxiliary variables. The goal of including the auxiliary variables is to account for differences in student characteristics, so that those factors do not contaminate the estimate of the effect of school type on test scores.

The originality of Leamer's contribution is that he articulated a previously under-emphasized conflict between the mathematical derivations of inferential statistics in regression (i.e., significance levels and confidence intervals) and the routine practice of applied econometricians. The inferential statistics, as reported by computer programs, are applicable to an experimental setting. But most economic research is conducted on data generated by uncontrolled economic activity, not by a controlled experiment. In practice, the econometrician tries a large number of regressions before reporting the ones that he or she thinks are most informative. Some econometricians, most notably David F. Hendry and his collaborators in the United Kingdom and elsewhere, embraced the use of specification searches, and developed progressively more complicated algorithms to automate them (see, e.g., Krolzig and Hendry 2001). Leamer took the opposite position on specification searches. He forced the econometric theorists to deal with the reality of econometric practice, and he forced applied researchers to face up to the corruption of their reported inferential statistics. Once the power of Leamer's critique sunk in, top journals became far less inclined to accept empirical work of the sort that had routinely been published before he wrote.

After Leamer pointed out problems inherent in the multiple-regression approach and the inevitable specification searching that it involves, economists have turned to quasi-experimental methods, as surveyed by Joshua Angrist and Jörn-Steffen Pischke (2010). To examine charter school performance, a researcher might look for instances in which admission to the charter school was determined by lottery. The researcher would compare students who attended the charter

school with public-school students who had entered the lottery but were not selected. The assumption would be that there are no systematic differences between students who won the lottery and students who lost the lottery, other than the type of school that these students wound up attending.

The significance of what Angrist and Pischke (2010) termed the "credibility revolution in empirical economics" can be seen in the John Bates Clark Medal awards given to researchers who participated in that revolution. Between 1995 and 2015, of the fourteen Clark Medal winners, by my estimate at least seven—David Card, Steven Levitt, Esther Duflo, Amy Finkelstein, Raj Chetty, Matthew Gentzkow, and Roland Fryer—are known for their empirical work using research designs intended to avoid the regression problems that Leamer highlighted. Duflo shared the Nobel Prize in 2019 for work that, atypically in economics, involved conducting field experiments rather than finding quasi-experiments in existing data.

Leamer has somewhat different views for how best to proceed with non-experimental data. Indeed, he has criticized some of the currently popular research methods (Leamer 2010). In his book *Macroeconomic Patterns and Stories* (Leamer 2009) he demonstrates an idiosyncratic approach to the challenges of finding convincing, robust patterns in macroeconomic data.

Here I explain Leamer's critique, as spelled out in his 1978 book and his 1983 AER article. I review Leamer's 2009 book on empirical macroeconomics, and I summarize his doubts about the 'credibility revolution.'

Leamer's long career also includes research in international trade and in regional economic forecasting. I omit that work from this essay.

Specification searches

The issue that drew Leamer's attention was this: In theory the multiple-regression approach means using the computer to estimate the one equation that the investigator knows best characterizes the data; in practice however investigators do not start out knowing the best way to characterize the data and they undertake an iterative, trial-and-error process in which they use the computer to estimate many different equations utilizing the same data. The trials in such a process are what Leamer (1978) termed "specification searches."

Econometricians use specification searches because they did not conduct the process that generated their data. Therefore, they need to explore their data in order to learn more about the behavior involved. As Leamer writes, "an experiment defines a model. When a specification search occurs, the researcher reveals that he does not think an experiment was conducted" (Leamer 1978, 2).

With an experiment, the investigator controls how the data are generated.

The statistical analysis can be based on that a priori knowledge. With observational data—meaning data that are generated in the world by a process that the investigator does not control—the investigator is uncertain about what model should be used to describe the data. The iterative, trial-and-error process is used to discover the model that the investigator believes is most appropriate.

Leamer argues that with observational data specification searches are inevitable and necessary. But from the standpoint of statistical theory, what practitioners do is simply bad. They violate what Leamer (1978, 4) terms "the Axiom of Correct Specification."

> The Axiom of Correct Specification
>
> (a) The set of explanatory variables that are thought to determine (linearly) the dependent variable must be
>
> (1) unique,
>
> (2) complete,
>
> (3) small in number, and
>
> (4) observable.
>
> (b) Other determinants of the dependent variable must have a probability distribution with at most a few unknown parameters.
>
> (c) All unknown parameters must be constant.
>
> If this axiom were, in fact, accepted, we would find one equation estimated for every phenomenon… Quite the contrary, we are literally deluged with regression equations, all offering to "explain" the same event…
>
> "Believers" use ad hoc techniques to search for specification, throwing out insignificant variables here and there, for example, but they continue to regard the end result of such a methodology to be identical to the end result obtained in the experimental sciences (or at least cynically to act that way). Believers report the summary statistics from the nth equation as if the other $n - 1$ were not tried, as if the nth equation defined a controlled experiment.
>
> At the other extreme are agnostics, who gladly admit the irrelevance of classical inference. … Agnostics may discount any statistical result until it has been employed in a prediction outside the data period. (Leamer 1978, 4)

Leamer argues that the practicing econometrician's exploration of alternative specifications is a proper adaptation to the situation one confronts in a non-experimental setting. Given that you did not design the process that generated the data, you *must* engage in exploratory analysis. Otherwise, you are likely to overlook important information. In his AER paper, Leamer writes

> A study of the anomalies of the data is what I have called "Sherlock Holmes" inference, since Holmes turns statistical inference on its head: "It is a capital mistake to theorize before you have all the evidence. It biases the judgements." (Leamer 1983, 40)

In *Specification Searches*, Leamer writes

> Sherlock solves the case by weaving together all the bits of evidence into a plausible story. He would think it indeed preposterous if anyone suggested that he should construct a function indicating the probability of the particular evidence at hand for all possible hypotheses and then assign prior probabilities to the hypotheses… Sherlock Holmes procedures are an essential feature of scientific learning. But when models are instigated by the data, the traditional theories of inference are, regrettably, invalidated. (Leamer 1978, 286–287)

The Sherlock Holmes method of using the data to generate hypotheses is a key motivation for specification searches. It is the subject of an entire chapter of Leamer's book and is the main form of specification search discussed in the AER piece. In the article, Leamer goes on to write:

> [A] theory constructed before seeing the facts can be disastrously inappropriate and psychologically difficult to discard. But if theories are constructed after having studied the data, it is difficult to establish by how much, if at all, the data favor the data-instigated hypothesis. For example, suppose I think that a certain coefficient ought to be positive, and my reaction to the anomalous result of a negative estimate is to find another variable to include in the equation so that the estimate is positive. Have I found evidence that the coefficient is positive? (Leamer 1983, 40)

Consider the hypothetical example of comparing charter schools with public schools. Suppose that in the first regression that we run, controlling for the income and educational attainment of parents, we find a small, insignificant advantage for public schools. We are disappointed by this result, and we explore the data further. We consider that ethnicity may have a causal effect that is not picked up by parents' income and education. We add to the equation variables that will indicate whether a student is African-American or Hispanic or Asian. In this equation, charter schools deliver significantly higher test scores, and this is the result that we report.

From the perspective of the Axiom of Correct Specification, we should never have tried the second regression. But it is neither realistic nor wise to suggest that only one equation can be tried with the data. What Leamer points out is that the specification search is not wrong in itself. What's misleading is the *reporting of the results*. Rather than proposing to end the practice of specification searches, Leamer tries to come up with better ways to undertake reporting of the process. As we will see, the economics profession took the problem posed by specification searching seriously, but it gravitated toward using different research designs rather than different reporting methods.

One approach that allows specification searches and still provides for valid

statistical inference is to randomly divide the data into two (not necessarily equal-size) samples. The investigation sample is used multiple times for trial-and-error specification searches, arriving at the investigator's preferred model. The second sample is 'held back' and used only once, to provide a statistically valid assessment of the preferred model. The results that the investigator reports would be based entirely on the results from the 'holdback sample.' This approach is most readily applied in situations involving 'big data' and machine learning, in which where there is sufficient data to provide both a large investigation sample and a large holdback sample. It has been scarcely practiced in economic research. It is not likely that there would be enough data available to apply this method to the problem of comparing test scores at charter schools with those at public schools.

In *Specification Searches*, Leamer suggested a different approach in a Bayesian spirit. He suggested that each time you try a new specification, you should give weight to the previous specification in the sense of a Bayesian prior. If I first looked at test scores across schools without controlling for ethnicity, then I behaved as if I had a prior belief that the coefficients on these variables were likely to be near zero. Subsequently, when I include these variables, I should use a Bayesian prior distribution that incorporates that prior belief. The result will be coefficients somewhere in between what the data find and the prior of zero.

But such an approach involves many practical problems. For one thing, two different researchers, using the same data, could arrive at different coefficients on variables based on the order in which equations are estimated. Suppose I include only X in my first specification and you include only Z in your first specification, but then both of us estimate a second equation that includes both X and Z. When we report coefficients, my coefficient on X will be larger than yours and my coefficient on Z will be smaller than yours.

In later writing Leamer advocated instead for a more conscientious approach, which might be termed comprehensive sensitivity analysis. It was common practice for economists to undertake ad hoc sensitivity analysis. The investigator might have attempted, say, 100 different specifications, and then selected three or four of these to report in the published paper, implicitly claiming that these reported specifications cover the range of reasonable inferences from the data. In his AER paper, Leamer argued that this type of sensitivity analysis is inadequate.

> The defect of this style is that the coverage of assumptions is infinitesimal, in fact a zero volume set in the space of assumptions. What is needed instead is a more complete, but still economical way to report the mapping of assumptions into inferences. (Leamer 1983, 38)

Leamer proposed a method, sometimes called "extreme bounds analysis," in which one should attempt every plausible specification and then report for the parameter of interest the range of consequent values. If the range is narrow, then inference about the parameter is robust. If not, then any inference is fragile. He described it thusly:

> Include in the equation the treatment variable and a single linear combination of the additive controls. Then find the linear combination of controls that provides the greatest estimated treatment effect and the linear combination that provides the smallest estimated treatment effect. That corresponds to the range of estimates that can be obtained when it is known that the controls are doubtful. (Leamer 2010, 37)

Leamer's recommendation was criticized by Michael McAleer, Adrian Pagan, and Paul Anthony Volker (1985) and was not widely adopted by practitioners. Instead, what eventually transpired was what Angrist and Pischke called the 'credibility revolution.' Economists looked for quasi-experimental circumstances embedded in observational data. Recall our example of comparing students who won the lottery to get into charter schools with students who lost such a lottery. If the lottery is truly random, then there should be no systematic differences between winners and losers in terms of parental income, parental educational attainment, ethnicity, or other variables that might affect outcomes. The investigator can justify omitting all such variables from the study, thus avoiding the problem of specification searches. In less ideal situations where assignment is random conditional on controls, the selection of such controls is at least limited to those that are plausibly related to assignment, mitigating the impact of search.

In effect, much of the profession decided that arguments about empirical values could not be settled by any method that included specification searches. Specification searches create too many opportunities for the investigator to present unreliable results, either intentionally or accidentally, through choice of specification. In practice the parameters estimated using these conventional regression techniques on observational data were likely to be fragile, whether or not they were reported as such. Economists have come to express more confidence in the robustness of estimates obtained from quasi-experimental methods.

Leamer the macroeconomist

"The bad news is that I am not a macroeconomist and thus cannot claim an expert's knowledge of the theory of the field. The good news is that I am not a macroeconomist, and I do not carry the heavy intellectual baggage that most

macroeconomists lug around." So writes Leamer in the preface to *Macroeconomic Patterns and Stories* (2009, v–vi). That book was positioned as a textbook—it was subtitled "A Guide for MBAs." But in fact it is a treatise that displays an original macroeconomic perspective that should have been a strong candidate for replacing conventional approaches in both teaching and research. The book deserves to be studied and its ideas debated within the highest reaches of the macroeconomic profession.

Leamer opens with a powerful epistemological statement:

> You may want to substitute the more familiar *scientific* words "theory and evidence" for "patterns and stories." Do not do that. With the phrase "theory and evidence" come hidden stow-away after-the-fact myths about how we learn and how much we can learn. The words "theory and evidence" suggest an incessant march toward a level of scientific certitude that cannot be attained in the study of the complex, self-organizing human system that we call the economy. The words "patterns and stories" much more accurately convey our level of knowledge, now, and in the future as well. It is literature, not science. (Leamer 2009, 3)

Perhaps it would have been better to equate macroeconomic pattern-seeking and storytelling with the study of history. As with history, macroeconomic events take place in sequence. We have only one sequence of events to study, with many possible explanatory variables.

The difficulty with macroeconometrics became apparent to me when I first read *Specification Searches*. This was in 1982, at a time when empirical macroeconomists were just coming to grips with problems caused by very high serial correlation. Because macroeconomic time series tend to have high serial correlation, when they are measured in levels all variables appear to be highly correlated. But when the serial correlation is treated by measuring the variables in terms of change from quarter to quarter, the correlation often drops to near zero. Equations uncorrected for serial correlation are severely biased. Equations properly corrected tend to be uninformative.[2] Either way, coefficients in structural macroeconometric models are fragile in the extreme.

The 'credibility revolution' largely bypassed macroeconomics. Angrist and Pischke (2010, 6) wrote that in empirical macroeconomics "progress—by our lights—is less dramatic." In my view, this is an understatement. In fact, no one has been able to revive the project of estimating structural macroeconometric models (with a consumption function, an investment function, a Phillips Curve, a money-demand function, and so on) using quasi-experimental methods. Leamer writes:

2. I further discuss the problems with empirical macroeconomics in Kling (2011; 2012).

> Finally, I think that Angrist and Pischke are way too optimistic about the prospects for an experimental approach to macroeconomics. Our understanding of causal effects in macroeconomics is virtually nil, and will remain so. (Leamer 2010, 44)

Conventional macroeconomists deal in equations, such as a Phillips Curve or a money-demand function. Because macroeconomic data are quantitative, expressing macroeconomic theory in terms of equations may seem like a plausible approach. But because macroeconomic data reflect a single historical unfolding of events, the use of equations is not really justified. Just as a historian seeking to understand the rise and fall of empires or the outcomes of wars would not hope to gain much insight by writing down a mathematical model with a few equations, macroeconomists who bury themselves in equations are mostly wasting their time.

In *Macroeconomic Patterns and Stories*, Leamer eschews the usual approach of treating the economy as a system of equations. Instead, he examines the patterns in the data and then invites the reader to consider what sorts of stories might explain those patterns.

For each important macroeconomic variable, Leamer first examines how it is measured. He explains how statistical agencies arrive at numbers for Gross Domestic Product, employment, inflation, and so on. He then plots the data in terms of levels, quarterly changes, and moving averages of quarterly changes. Finally, he plots the data relative to quarters that the National Bureau of Economic Research has decided by committee vote were cyclical peaks.

At each step in his analysis, Leamer finds phenomena that are often overlooked by even the leading practitioners of macroeconomics. For example, macroeconomic data is routinely adjusted for regular seasonal fluctuations, and these adjustments are large. The unadjusted numbers show a drop from the pre-Christmas to the post-Christmas season that is much steeper than we see during bad recessions. As Leamer puts it:

> Though on average, nominal GDP has grown at the rate of 9.2 percent, the January–March first-quarter production has been down compared with the fourth quarter at the annualized rate of about −20%! After working so hard to get those gifts ready for the holiday, Santa and his elves huddle around the fireplaces in the winter and do not get much done. (Leamer 2009, 37–38)

Leamer also notes that the extent of seasonality has been declining over time. The decline points to another challenge overlooked by many macroeconomists: the structure of the economy keeps evolving. The share of workers employed in service industries has increased relative to that in goods-producing industries. The use of computers for tracking and forecasting has reduced the magnitude of inventory

fluctuations. General price inflation has become less variable cyclically, particularly when compared with secular divergence involving large price increases in higher education and health care with declining prices in computers and communications.

Looking at data on changes in employment using both monthly changes and three-month moving averages of monthly changes, Leamer discovered a phenomenon that to my knowledge no one else had previously found: momentum. That is, when the three-month increase in payrolls is well below average, the following month's increase is very likely also to be well below average. But for this momentum to become established, it takes three months of below-average employment growth.

Another device Leamer uses is a chart of recessions and recoveries relative to the dating established by the NBER cyclical dating committee. This plot

> allows a direct side-by-side comparison of the ten recessions since 1948. The horizontal axis is the number of quarters before or after the NBER peak. To the left are quarters before the recessions begin; to the right are recession quarters. What is displayed is the percentage difference between employment and the employment level at the peak. (Leamer 2009, 111–112)

What one sees in these data is that the early postwar fluctuations were short and sharp. Employment fell dramatically during recessions but recovered quickly afterward. Periods of growth were interrupted relatively soon by another recession.

More recently, growth episodes have been milder but longer lasting. This pattern also holds for the period following the severe recession of 2008–2009, which took place subsequent to the appearance of *Macroeconomic Patterns and Stories*. To me, this suggests that the phenomenon of recession and recovery is now very different from what it was decades ago. Recessions used to be dominated by manufacturing layoffs, with workers returning quickly to their previous jobs on the assembly line once excess inventories had been worked off. In the more recent recessions, inventory corrections have not played an important role. Workers have lost jobs to which they were never going to return, and the process of creating new jobs during a recovery has proceeded relatively slowly.

The goal of Leamer's extensive empirical analysis is not to arrive at a set of equations to describe the economy. Instead, he arrives at stories. One of his key chapters is entitled "Idleness Stories," in which he tries to explain how it is that the processes of market adjustment fail to ensure continuous full employment. He notes that employment is not just a momentary state. It is a long-term relationship.

> This gives us a good unemployment story. In recessions with limited cash flows, facing difficulties servicing debts and making payrolls, some employers simply go out of business, ending forever their relationships with employees.

> Other employers may continue to operate, but reluctantly sever their relationships with some employees. (Leamer 2009, 146)

In the subsequent chapter, Leamer tells "Cycle Stories," attempting to explain how fluctuations in business conditions occur. His stories are based on the fact that the greatest cyclical volatility is in manufacturing and construction.

Overall, Leamer demonstrates the virtues of Sherlock Holmes analysis, of carefully examining the data in order to develop hypotheses. As a result, his exploration opens up different perspectives not found in conventional macroeconomics. If economists would read *Macroeconomic Patterns and Stories* with an open mind, without commitment to conventional theoretical frameworks, I am convinced that they would come away with many new insights. In my case, reading the book influenced me to think much more in terms of structural variation in the economy and to think less in terms of treating the economy as if it were one big factory producing GDP.

Leamer on the shortcomings of natural experiments

Leamer (2010) responded to the essay in which Angrist and Pischke trumpeted the 'credibility revolution' of using quasi-experimental methods rather than specification searches.

> As Angrist and Pischke persuasively argue, either purposefully randomized experiments or accidentally randomized "natural" experiments can be extremely helpful, but Angrist and Pischke seem to me to overstate the potential benefits of the approach. (Leamer 2010, 32)

One problem Leamer notes is the issue of whether results found in a narrow quasi-experimental setting can be generalized to other contexts.

> For example, how does [David] Card's (1990) study of the effect on the Miami labor market of the Mariel boatlift of 125,000 Cuban refugees in 1980 inform us of the effects of a 2000 mile wall along the southern border of the United States? (Leamer 2010, 33)

Another example of the problem of extrapolating findings in one setting to a different context that Leamer points to is the behavior of mortgage-related securities during the financial crisis. Bond rating agencies and other quantitative analysts made predictions about the performance of these securities based on

historical norms for housing markets that did not apply in 2006–2008.

Because of the extrapolation problem, Leamer writes:

> I thus stand by the view in my 1983 essay that econometric theory promises more than it can deliver, because it requires a complete commitment to assumptions that are actually only half-heartedly maintained. The only way to create credible inferences with doubtful assumptions is to perform a sensitivity analysis that separates the fragile inferences from the sturdy ones: those that depend substantially on the doubtful assumptions and those that do not. Since I wrote my "con in econometrics" challenge much progress has been made in economic theory and in econometric theory and in experimental design, but there has been little progress technically or procedurally on this subject of sensitivity analyses in econometrics. Most authors still support their conclusions with the results implied by several models, and they leave the rest of us wondering how hard they had to work to find their favorite outcomes and how sure we have to be about the instrumental variables assumptions with accidentally randomized treatments and about the extent of the experimental bias with purposefully randomized treatments. It's like a court of law in which we hear only the experts on the plaintiff's side, but are wise enough to know that there are abundant arguments for the defense. (Leamer 2010, 36)

Today, we know that the 'credibility revolution' was followed a few years later by the 'replication crisis.' That would seem to vindicate Leamer.

Friedrich Hayek titled his 1974 lecture accepting the Nobel Prize "The Pretence of Knowledge." He wrote:

> Unlike the position that exists in the physical sciences, in economics and other disciplines that deal with essentially complex phenomena, the aspects of the events to be accounted for about which we can get quantitative data are necessarily limited and may not include the important ones. While in the physical sciences it is generally assumed, probably with good reason, that any important factor which determines the observed events will itself be directly observable and measurable, in the study of such complex phenomena as the market, which depend on the actions of many individuals, all the circumstances which will determine the outcome of a process, for reasons which I shall explain later, will hardly ever be fully known or measurable. (Hayek 1992/1974)

Along similar lines, Leamer writes:

> Let's face it. The evolving, innovating, self-organizing, self-healing human system we call the economy is not well described by a fictional "data-generating process." The point of the sensitivity analyses that I have been

> advocating begins with the admission that the historical data are compatible with countless alternative data-generating models. If there is one, the best we can do is to get close; we are never going to know it. (Leamer 2010, 38)

Edward Leamer emphasizes to economists the uncertainty that we face and the humility with which we should make our claims. The profession has benefited from partially listening to Leamer concerning the weaknesses of multiple-regression methods. It would benefit even more by taking all of his methodological suggestions to heart.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics. *Journal of Economic Perspectives* 24(2): 3–30. **Link**

Begley, C. Glenn, and John P. A. Ioannidis. 2015. Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circulation Research* 116(1): 116–126.

Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, et al. 2018. Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* Between 2010 and 2015. *Nature Human Behaviour* 2: 637–644.

Hayek, Friedrich A. 1992 [1974]. The Pretence of Knowledge. In *Nobel Lectures, Economics 1969–1980*, ed. Assar Lindbeck. Singapore: World Scientific. **Link**

Ioannidis, John P. A. 2005. Why Most Published Research Findings Are False. *PloS Medicine* 2(8): 696–701. **Link**

Ioannidis, John P. A. 2016. Evidence-Based Medicine Has Been Hijacked: A Report to David Sackett. *Journal of Clinical Epidemiology* 73: 82–86.

Kling, Arnold. 2011. Macroeconometrics: The Science of Hubris. *Critical Review* 23(1–2): 123–133.

Kling, Arnold. 2012. Patterns of Sustainable Specialization and Trade: A Smith-Ricardo Theory of Macroeconomics. Adam Smith Institute (London). **Link**

Krolzig, Hans-Martin, and David F. Hendry. 2001. Computer Automation of General-to-Specific Model Selection Procedures. *Journal of Economic Dynamics and Control* 25: 831–866.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Hoboken, N.J.: Wiley.

Leamer, Edward E. 1983. Let's Take the Con Out of Econometrics. *American Economic Review* 73(1): 31–43.

Leamer, Edward E. 2009. *Macroeconomic Patterns and Stories: A Guide for MBAs*. New York: Springer.

Leamer, Edward E. 2010. Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives* 24(2): 31–46. **Link**

McAleer, Michael, Adrian R. Pagan, and Paul A. Volker. 1985. What Will Take the Con Out of Econometrics? *American Economic Review* 75(3): 293–307.

About the Author

Arnold Kling is a member of the Financial Markets Working Group of the Mercatus Center at George Mason University. He is the author of seven books, most recently *The Three Languages of Politics* and *Memoirs of a Would-Be Macroeconomist*. He writes at Askblog (**link**). His email is arnold@arnoldkling.com.

Go to archive of Watchpad section
Go to March 2020 issue