

Wallowing in significance

The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. By Stephen T. Ziliak and Deirdre N. McCloskey, Ann Arbor, MI, The University of Michigan Press, 2007, 321 pp., \$24.95/paper.

All but the most astute BLS news release reader might overlook the note near the end of some BLS regional reports that states, “A value that is statistically different from another does not necessarily mean that the difference has economic or practical significance.” This reviewer thought he understood what that meant, but after a reading of *The Cult of Statistical Significance*, that statement has taken on new “significance.”

Stephen T. Ziliak and Deirdre N. McCloskey are “not professional statisticians, only amateur historians and philosophers of science.” They are both professors and economists who are also artful writers. Ziliak has taught at Emory University and the Georgia Institute of Technology; he is currently Professor of Economics at Roosevelt University. Ziliak’s resume includes a stint working as a state labor market analyst, in which he was not able to provide black teenage unemployment rates because they did not meet an “arbitrary level of statistical significance.” McCloskey, the Distinguished Professor of Economics, History, English, and Communication at the University of Illinois at Chicago, has authored 20 books and 300 articles. This reviewer was first introduced to McCloskey’s work over 20 years ago when a colleague shared the article, “Economic Writing” (Western Economic Association, *Economic Inquiry*, April 1985), an entertaining and engaging piece that provides writing guidance

to economists.

According to *Cult’s* authors, the problem is that significance has become a broken, or highly overused and abused, statistical instrument. “The offering of statistically significant coefficients seems ceremonial,” write Professors Ziliak and McCloskey, who document a history of the problem while attacking its misuse. “In statistical fields such as economics...the idol is the test of significance.” Put succinctly, Ziliak and McCloskey feel statistical significance is simply bad science—“One erects little ‘significance’ hurdles, six inches tall, and makes a great show of leaping over them, concluding from a test of statistical significance that the data are ‘consistent with’ one’s own very charming hypothesis.”

Their point of contention is that “fit is not the same thing as importance. Statistical significance is not the same thing as scientific finding.” A scientific study is concerned with determining the magnitude of effect, answering the question, “How much?” Contrast this with conclusions based solely on a statement of statistical significance. The difference is one of what the authors call “oomph” versus a “philosophy of mere existence.” This point is masterfully illustrated with a number of case histories (including the 1990 South Carolina salmonella outbreak and studies on both St. John’s-Wort and Vioxx).

In 1996, the authors analyzed scholarly American Economic Review (AER) articles from the 1980s, subjecting them to 19 critical evaluative questions, in order to assess the quality of their statistical analyses. Among their findings was that 70 percent of the applied econometric papers published made no distinction between statistical significance and economic significance. The authors

repeated the study with articles from the 1990s, and the results were not much better.

“No competent statistician would recommend,” write Ziliak and McCloskey, “that economists use only tests of statistical significance without a loss function or a consideration of power...” Explain the *Cult’s* authors, “Power asks, ‘What in the proffered experiment is the probability of correctly rejecting the null hypothesis, concluding that the null hypothesis is indeed false when it is false?’” Ziliak and McCloskey assert, “Calculations of Type I error pretend otherwise... they act as if the null hypothesis... is the only hypothesis that is worthy of probabilistic assessment. They ignore the other hypotheses.”

To help solve the statistical significance problem, Ziliak and McCloskey propose issuing a “Statement on the proprieties of substantive significance” and distributing it to editors and researchers. “Undergraduates need to hear from the beginning that size matters,” state the authors. Size matters from more than one perspective: in terms of the size of the error (and, the authors point out, random error is but “one out of many dozens of errors and seldom the biggest”); in terms of sample size; and in terms of the size of the observed economic effect.

How did it happen that statistical significance became the expected and most abused litmus test of modern research? McCloskey and Ziliak raise a number of possibilities, including sociological reasons, to explain the current situation. “Testimators rest content with a nominal level of statistical significance, ignoring the real significance—the rise or fall in the price of the ostensible object of inquiry. Suffering from precision illusion, they ignore real error.”

In addition to exposing us to the

development of ideas, the authors also paint a picture of the personalities behind the number theories. This added color, though sometimes entertaining, may occasionally border on character attack. Some of the portrayals, in this reviewer's opinion, may have detracted from the book's potency.

Nevertheless, the message remains: Even employees of major U.S. statistical agencies might take statistical significance for granted. After all, we and other statistical practitioners and data disseminators know all about estimate formulation and sample error. We can analyze data and present

our survey results and research findings to the public, providing valuable information about our economy. Relatively few of us, however, know the history of significance analysis, the controversy that surrounds its use, and the "substantive" strength added by considerations of power and other analytical methods.

Cult's strength is that it fills that void...and then some. The authors are not shy about their message: "We hope you, oh significance tester, will read the book optimistically—with a sense of how "real" significance can transform your science." Whether or not one agrees with their conclusions,

some benefit might still accrue from a close reading of this work. Beyond the many-faceted descriptions of the problem, *Cult* provides a "reader's guide" for further direction and additional background in statistical testing estimation and error. And, if you are a researcher, the most valuable part of this work might be the discussion that surrounds Ziliak and McCloskey's 19-question AER evaluation—how would *your* study fare? □

—Bruce Bergman
New York Office
Bureau of Labor Statistics

Book review interest?

Interested in reviewing a book for the *Monthly Labor Review*? We have a number of books by distinguished authors on economics, industrial relations, other social sciences, and related issues waiting to be reviewed. If you have good writing skills and/or experience, then please contact us via E-mail at mlr@bls.gov.