The Journal of Finance / Volume 72, Issue 4 / p. 1399-1440

ARTICLE

Presidential Address: The Scientific Outlook in Financial Economics

CAMPBELL R. HARVEY

First published: 08 July 2017
https://doi.org/10.1111/jofi.12530
Citations: 372

Abstract

ABSTRACT

Given the competition for top journal space, there is an incentive to produce "significant" results. With the combination of unreported tests, lack of adjustment for multiple tests, and direct and indirect *p*-hacking, many of the results being published will fail to hold up in the future. In addition, there are basic issues with the interpretation of statistical significance. Increasing thresholds may be necessary, but still may not be sufficient: if the effect being studied is rare, even $t > 3$ will produce a large number of false positives. Here I explore the meaning and limitations of a *p*-value. I offer a simple alternative (the minimum Bayes factor). I present guidelines for a robust, transparent research culture in financial economics. Finally, I offer some thoughts on the importance of risk-taking (from the perspective of authors and editors) to advance our field.

SUMMARY

- Empirical research in financial economics relies too much on *p*-values, which are poorly understood in the first place.
- Journals want to publish papers with positive results and this incentivizes researchers to engage in data mining and "*p*-hacking."
- The outcome will likely be an embarrassing number of false positives—effects that will not be repeated in the future.
- The minimum Bayes factor (which is a function of the *p*-value) combined with prior odds provides a simple solution that can be reported alongside the usual *p*-value.
- The Bayesianized *p*-value answers the question: What is the probability that the null is true?
- The same technique can be used to answer: What threshold of *t*-statistic do I need so that there is only a 5% chance that the null is true?
- The threshold depends on the economic plausibility of the hypothesis.

REFERENCES

Alter, Adam L., and Daniel M. Oppenheimer, 2006, Predicting short-term stock fluctuations by using processing fluency, *Proceedings of the National Academy of Sciences* **103**, 9369–9372.

CAS | PubMed | Web of Science® | Google Scholar

American Finance Association, 2016, Code of Professional Conduct and Ethics. Available at http://www.afajof.org.

Google Scholar

American Statistical Association, 2016, Statement on Statistical Significance and P-Values. Available at: https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf.

Google Scholar

Bayarri, M. J., Daniel J. Benjamin, James O. Berger, and Thomas M. Sellke, 2016, Rejection odds and rejection ratios: A proposal for statistical practice in testing hypotheses, *Journal of Mathematical Psychology* **72**, 90–103.

CAS PubMed Web of Science® Google Scholar

Bayarri, M. J., and James O. Berger, 1998, Quantifying surprise in the data and model verification, in José M. Bernardo, James O. Berger, A. Philip Dawid, and Adrian F. M. Smith, eds.: *Bayesian Statistics*. Vol. **6**. (Oxford University Press, Oxford).

Google Scholar

Bem, Daryl J., 2011, Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect, *Journal of Personality and Social Psychology* **100**, 407–425.

PubMed Web of Science® Google Scholar

Berger, James O., and Thomas Sellke, 1987, Testing a point null hypothesis: The irreconcilability of *p* values and evidence, *Journal of the American Statistical Association* **82**, 112–122.

Web of Science® Google Scholar

Berk, Jonathan, Campbell R. Harvey, and David A. Hirshleifer, 2017, How to write an effective referee report and improve the scientific review process, *Journal of Economic Perspectives* **31**, 231–244.

Web of Science® Google Scholar

Berkson, Joseph, 1938, Some difficulties of interpretation encountered in the application of the chi-square test, *Journal of the American Statistical Association* **33**, 526–536.

Web of Science® Google Scholar

Berkson, Joseph, 1942, Tests of significance considered as evidence, *Journal of the American Statistical Association* **37**, 325–335.

Web of Science® Google Scholar

Cacioppo, John T., Stephanie Cacioppo, Gian C. Gonzaga, Elizabeth L. Ogburn, and Tyler J. VanderWeele, 2013, Marital satisfaction and break-ups differ across on-line and off-line meeting venues, *Proceedings of the National Academy of Sciences* **110**, 10135–10140.

CAS | PubMed | Web of Science® | Google Scholar

Carver, Ronald P., 1978, The case against statistical significance testing, *Harvard Educational Review* **48**, 378–399.

Web of Science® | Google Scholar

CERN, 2017, LHC: The guide, Conseil Européen pour la Recherche Nucléaire, January. Available at http://cds.cern.ch/record/2255762/files/CERN-Brochure-2017-002-Eng.pdf.

Google Scholar

Churchill, Gary A., 2014, When are results too good to be true? *Genetics* **198**, 447–448.

PubMed | Web of Science® | Google Scholar

Cohen, Jacob, 1994, The earth is round (p<.05), *American Psychologist* **49**, 997–1003.

Web of Science® | Google Scholar

Coles, Jeffrey, Naveen Daniel, and Lalitha Naveen, 2008, Boards: Does one size fit all? *Journal of Financial Economics* **87**, 329–356.

Web of Science® | Google Scholar

Comte, Auguste, 1856, *The Positive Philosophy of Auguste Comte*, translated by Harriett Marineau (Calvin Blanchard, New York). Vol. **II**. Available at: https://doi.org/10.1017/CBO9780511701467.

Google Scholar

Consolidated Standards of Reporting Trials, 2010. Available at http://www.consort-statement.org/consort-2010.

Google Scholar

Cook, John, and Stephan Lewandowsky, 2012, *The Debunking Handbook* (University of Queensland, St. Lucia, Australia). Available at: http://www.skepticalscience.com/docs/Debunking_Handbook.pdf.

Google Scholar

Edwards, Ward, Harold Lindman, and Leonard J. Savage, 1963, Bayesian statistical inference for psychological research, *Psychological Review* **70**, 193–242.

Web of Science® | Google Scholar

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* **47**, 427–465.

Web of Science® | Google Scholar

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* **116**, 1–22.

Web of Science® | Google Scholar

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* **81**, 607–636.

Web of Science® | Google Scholar

Fanelli, Daniele, 2010, ''Positive'' results increase down the Hierarchy of the Sciences, *PLoS ONE* **5**, e10068.

CAS | PubMed | Web of Science® | Google Scholar

Fanelli, Daniele, 2012, Negative results are disappearing from most disciplines and countries, *Scientometrics* **90**, 891–904.

Web of Science® | Google Scholar

Fanelli, Daniele, 2013, Positive results receive more citations, but only in some disciplines, *Scientometrics* **94**, 701–709.

Web of Science® | Google Scholar

Fisher, Ronald A., 1925, *Statistical Methods for Research Workers* (Oliver and Boyd Ltd., Edinburgh).

Web of Science® | Google Scholar

Francis, G., 2012, Too good to be true: Publication bias in two prominent studies from experimental psychology, *Psychonomic Bulletin and Review* **19**, 151–156.

PubMed | Web of Science® | Google Scholar

Gauchat, Gordon, 2012, Politicization of science in the public sphere: A study of public trust in the United States, 1974–2010, *American Sociological Review* **77**, 167–187.

Web of Science® | Google Scholar

Gawande, Atul, 2016, The mistrust of science, Commencement speech, California Institute of Technology, reprinted in *The New Yorker*. Available at http://www.newyorker.com/news/news-desk/the-mistrust-of-science.

Google Scholar

Gelman, Andrew, and Eric Loken, 2013, The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time, Working paper, Columbia University. Available at: http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Google Scholar

Gelman, Andrew, and Eric Loken, 2014, The statistical crisis in science, *American Scientist* **102**, 460–465.

Web of Science® | Google Scholar

Gelman, Andrew, and Hal Stern, 2006, The difference between "significant" and "not significant" is not itself statistically significant, *The American Statistician* **60**, 328–331.

Web of Science® | Google Scholar

Gigerenzer, Gerd, 2004, Mindless statistics, *Journal of Socio-Economics* **33**, 587–606.

Google Scholar

Goodman, Steven N., 1993, P values, hypothesis tests, and likelihood: Implications for epidemiology of a neglected historical debate, *American Journal of Epidemiology* **137**, 485–496.

CAS | PubMed | Web of Science® | Google Scholar

Goodman, Steven N., 1999, Toward evidence-based medical statistics. 1: The *p*-value fallacy, *Annals of Internal Medicine* **130**, 995–1004.

CAS | PubMed | Web of Science® | Google Scholar

Goodman, Steven N., 2001, Of *p*-values and Bayes: A modest proposal, *Epidemiology* **12**, 295–297.

CAS | PubMed | Web of Science® | Google Scholar

Greenhouse, J. B., 2012, On becoming a Bayesian: Early correspondences between J. Cornfield and L. J. Savage, *Statistics in Medicine* **31**, 2782–2790.

PubMed | Web of Science® | Google Scholar

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies* **29**, 5–68.

Web of Science® | Google Scholar

Head, Alex, Gary Smith, and Julia Watson, 2009, Would a stock by any other ticker smell as sweet? *Quarterly Review of Economics and Finance* **49**, 551–561.

Google Scholar

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions, 2015, The extent and consequences of *p*-hacking in science. *PLoS Biol* **13**, e1002106.

PubMed | Web of Science® | Google Scholar

Hubble, Edwin, 1954, *The Nature of Science and Other Lectures* (Huntington Library, San Marino, California).

Google Scholar

Hutton, Jane, 2010, Misleading statistics: The problems surrounding number needed to treat and number needed to harm, *Pharmaceutical Medicine* **24**, 145–149.

Google Scholar

Ioannidis, John P. A., 2005, Why most published research findings are false, *PLoS Medicine* **2**, e124.

PubMed | Web of Science® | Google Scholar

Ioannidis, John P. A., 2008, Why most discovered true associations are inflated (with discussion), *Epidemiology* **19**, 640–658.

PubMed | Web of Science® | Google Scholar

Jarosz, Andrew F., and Jennifer Wiley, 2014, What are the odds? A practical guide to computing and reporting Bayes factors, *Journal of Problem Solving* **7**, 2–9.

Google Scholar

Kerr, Norbert L., 1998, HARKing: Hypothesizing after the results are known, *Personality and Social Psychology Review* **2**, 196–217.

CAS | PubMed | Google Scholar

Lang, Janet M., Kenneth J. Rothman, and Cristina I. Cann, 1998, That confounded *p*-value (Editorial), *Epidemiology* **9**, 7–8.

CAS | PubMed | Web of Science® | Google Scholar

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* **47**, 13–37.

Web of Science® | Google Scholar

Marsh, David M., and Teresa J. Hanlon, 2007, Seeing what we want to see: Confirmation bias in animal behavior research, *Ethology* **113**, 1089–1098.

Web of Science® | Google Scholar

Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* **34**, 768–783.

Web of Science® | Google Scholar

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Web of Science® | Google Scholar

Neyman, Jerzy, and Egon S. Pearson, 1933, On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society* Series A **231**, 289–337.

Web of Science® | Google Scholar

Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl, 2012, Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability, *Perspectives on Psychological Science* **7**, 615–631.

PubMed | Web of Science® | Google Scholar

Nuzzo, Regina, 2014, Statistical errors, *Nature* **506**, 150–152.

CAS | PubMed | Web of Science® | Google Scholar

Ritchie, Stuart J., Richard Wiseman, and Christopher C. French, 2012, Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect, *PLoS ONE* **7**, e33423.

PubMed | Web of Science® | Google Scholar

Rosenthal, Robert, 1979, The "file drawer problem" and tolerance for null results, *Psychological Bulletin* **86**, 638–641.

PubMed | Web of Science® | Google Scholar

Royal Swedish Academy of Sciences, 2013, Press release, October 8. Available at **http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/press.html**.

Google Scholar

Russell, Bertrand, 1931, *The Scientific Outlook* (The Free Press, Glencoe, Illinois).

Google Scholar

Sever, Peter S., Björn Dahlöf, Neil R. Poulter, Hans Wedel, Gareth Beevers, Mark Caulfield, Rory Collins, Sverre E. Kjeldsen, Arni Kristinsson, Gordon T. McInnes, Jesper Mehlsen, Markku Nieminen, Eoin O'Brien, and Jan Östergren, 2003, Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have

average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid Lowering Arm (ASCOT-LLA): A multicentre randomised controlled trial, *The Lancet* **361**, 1149–1158.

CAS | PubMed | Web of Science® | Google Scholar

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19**, 425–442.

Web of Science® | Google Scholar

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons, 2014, *P*-curve: A key to the file-drawer, *Journal of Experimental Psychology: General* **143**, 534–547.

PubMed | Web of Science® | Google Scholar

Song, Fujian, A. J. Eastwood, Simon Gilbody, Lelia Duley, and A. J. Sutton, 2000, Publication and related biases, *Health Technology Assessment* **4**, 1–115.

PubMed | Web of Science® | Google Scholar

Srinivasan, Raji, and Nita Umashankar, 2014, There's something in a name: Value relevance of congruent ticker symbols, *Consumer Needs and Solutions* **1**, 241–252.

Google Scholar

Trafimow, David, and Michael Marks, 2015, Editorial, *Basic and Applied Social Psychology* **37**, 1–2.

Web of Science® | Google Scholar

Wasserstein, Ronald L., and Nicole A. Lazar, 2016, The ASA's statement on *p*-values: Context, process, and purpose, *The American Statistician* **70**, 129–133.

Web of Science® | Google Scholar

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**, 1455–1508.

Web of Science® | Google Scholar

Yermack, David, 1996, Higher market valuation of companies with a small board of directors, *Journal of Financial Economics* **40**, 185–212.

Web of Science® | Google Scholar

Citing Literature

Download PDF