REVIEW IN: STATISTICAL SOFTWARE NEWSLETTER IN COMPUTATIONAL STATISTICS + DATA ANALYSIS

DATE OF ISSUE: NOT GIVEN .... DATE RECEIVED: 22.10.04

Micah Altman, Jeff Gill & Michael P. McDonald (2004): *Numerical Issues in Statistical Computing for the Social Scientist.* John Wiley, Hoboken, NJ, xv + 323 pp., US $ 89.95, GB £ 52.95, € 75.00, ISBN 0-471-23633-0.

This is a guide to the pitfalls of statistical computing in the social sciences. The book is written by three political scientists, but most of the contents should also be of interest to researchers in sociology, psychology, or economics. As the authors state in the preface (p. xi), "[t]he overall purpose of this work is to address what we see as a serious deficiency in statistical work in the social and behavioral sciences, broadly defined. Quantitative researchers in these fields rely on statistical and mathematical computation as much as any of their colleagues in the natural sciences, yet there is less appreciation for the problems and issues in numerical computation." Quite recently, the problem has also attracted some attention in the econometrics literature (McCullough and Vinod, 1999). I suspect that to many social scientists it will be surprising that numerical computations to some extent depend on the specific software package and version, that some packages are more accurate than others, and that some are, in fact, not too useful overall.

Altman, Gill and McDonald begin with some juicy stories, including a startling example of naïve use of starting values published in *Science*. They proceed with fundamental concepts and terminology, such as "algorithms" vs. "heuristics" or "precision" vs. "accuracy", provide some background on computer arithmetic and the like, and discuss tools for the evaluation of statistical software. Most of the subsequent chapters, some of which are guest contributions, are devoted to specific types of problems that arise fairly often in social-science research: Markov chain Monte Carlo (ch. 5), numerical issues in connection with the inversion of Hessian matrices (ch. 6, by Jeff Gill and Gary King), ecological inference (ch. 7), nonlinear models (ch. 8, by Bruce McCullough), spatial econometrics (ch. 9, by James LeSage), and logistic regression (ch. 10, by Paul Allison). Although some of the problems appear to be genuine to political science (e.g., ecological inference), the bulk of the topics considered should be of much broader interest. For example, the chapter on logistic regression discusses the separation problem, a problem that, unfortunately, is still seldom discussed in textbooks written for social scientists, although it is not uncommon with small data sets.

In order to appreciate numerical problems and perhaps to be able to test one's favorite computing environment, an understanding of benchmark problems is useful. While no benchmarking method can prove that a software package is accurate, performing well on a benchmark data set provides some evidence that a piece of software is accurate for that domain of problems. Here the authors present a number of statistical benchmarks, including some nonlinear problems, tests of random number generators, and tests of distribution functions. A number of links to various resources including numerical libraries, benchmark data etc., are provided on a Web site accompanying the book, currently at `http://www.hmdc.harvard.edu/numerical_issues/`. The code and data sets used in the text are also available there.

The final chapter is particularly interesting in that it contains a number of recommendations for replication of quantitative research, a timely topic in statistics. In this regard, readers may also find the recent paper by Leisch and Rossini (2003) useful.

In summary, here we have a compact guide to the voluminous literature on optimization, numerical analysis, and computational statistics. This is no small achievement.

ISBN: 0471236330
Date of Publication: 20/01/04
Editor:
Mkt Manager: L. Cooksley.
Country:
Page 1 of 2 55.95

Leisch, F., and Rossini, A. (2003). Reproducible statistical research. *Chance*, 16, 46-50.

McCullough, B.D., and Vinod, H.D. (1999). The numerical reliability of econometric software. *Journal of Economic Literature*, 37, 633-665.

Christian Kleiber, Dortmund, Germany