

International Journal of Forecasting Guidelines for IJF Software Reviewers

It is desirable that there be some small degree of uniformity amongst the software reviews in this journal, so that regular readers of the journal can have some idea of what to expect when they read a software review. In particular, I wish to standardize the second section (after the introduction) of the review, and the penultimate section (before the conclusions). As stand-alone sections, they will not materially affect the reviewers ability to craft the review as he/she sees fit, while still providing consistency between reviews. This applies mostly to single-product reviews, but some of the ideas presented herein can be successfully adapted to a multi-product review.

The second section, *Overview*, is an overview of the package, and should include several things.

- Contact information for the developer, including website address.
- Platforms on which the package runs, and corresponding prices, if available.
- Ancillary programs included with the package, if any.
- The final part of this section should address Berk's (1987) list of criteria for evaluating statistical software. Relevant items from this list should be mentioned, as in my review of RATS (McCullough, 1997, pp.182-183).
- My use of Berk was extremely terse, and should be considered a lower bound. Feel free to amplify considerably, if the review warrants it. In fact, Berk's criteria, if considered in sufficient detail, could be the outline for a review itself.

The penultimate section, *Numerical Details*, directly addresses numerical accuracy and reliability, if these topics are not addressed elsewhere in the review. This section is not intended to be exhaustive, but merely to inform the reader of whether or not the developer pays attention to matters numerical.

- If relevant estimation benchmarks exist {it e.g.}, the StRD (McCullough, 1998, 1999a, 1999b) and Wilkinson's Tests (Sawitzki, 1994ab; McCullough, 2000),
 - does the developer provide benchmark results? If not, this important fact should be mentioned. By no means are you expected to benchmark the package. However, you are more than welcome to do so.
 - If the package offers full-information maximum likelihood, does it satisfy the Calzolari-Panattoni benchmark for Klein's Model I (Silk, 1998)? If not, invite the developer to say why not.
 - If the package offers GARCH(1,1), does it satisfy the Fiorentini-Calzolari-Panattoni (FCP) benchmark (McCullough and Renfro, 2000)? Again, invite the developer to say why not.
 - If the package has a random number generator, does the documentation give: the algorithm, its period, and the randomness tests it passes (McCullough, 1998)?
 - If the package has statistical distributions - e.g. for calculating p -values, etc.), does the documentation describe the bounds within which it is safe to use each one (McCullough, 1998)? For example, "The normal distribution provides six digits of accuracy for probabilities as small as $1E-12$."
 - Does the documentation describe the algorithms for linear regression? Is the Cholesky decomposition or the QR decomposition employed? Does the user have a choice?
 - What is the output from linear regression? Does it compute the Durn-Watson statistic when a lagged dependent is on the right-hand side? You don't have to mention what is presented, but you should note if any relevant statistics are not presented.
 - What about nonlinear least squares and maximum likelihood? What methods of solution are available? What options can the user invoke? What is the output from nonlinear procedures? Does it include the trace (some output, such as function value) at each iteration, or can an option be set to show the trace? Does it print out the gradient after convergence has been achieved? Does it show what options have been set?
 - Are partial correlation coefficients computed by the Yule-Walker equations (this is bad for economic data) or by least square or the Burg method (these are better)? See McCullough (1999c).
-
- Does the website post known bugs and/or bug-fixes? If not, this important fact should be mentioned.

When you are done with the review, please send me a copy (either pdf or postscript). If I have comments, I shall send them to you and ask for a rewrite. When I have signed off on the draft, please send a copy to the developer (with a CC to me) and ask them to send their comments/corrections to you with a CC to me. In that way, I can stay in the loop should any dispute arise.

You should correct any outright errors uncovered by the developer. You are not obliged to correct any statement of opinion with which the developer disagrees. If you encounter any difficulty with this, please do not hesitate to contact me.

If you encounter any bugs (some developers like to call them "undocumented features"), these should be noted in the "Numerical Details" section, if not noted elsewhere in the review.

Finally, in the introduction, please clearly identify your relationship with the package. Are you a regular user or just someone who happens to be doing a review. Please mention your qualifications for reviewing the package. If the package is about neural nets", for example, I certainly wouldn't commission a review by someone unfamiliar with the topic. So please self-cite a couple times, or otherwise indicate your familiarity with the topic.

References

Berk, K. (1987), Effective Microcomputer Statistical Software, *The American Statistician* 41, 222-228

McCullough, B. D. (1999). Algorithm Choice for (Partial) Autocorrelation Functions, *Journal of Economic and Social Measurement* 24, 265-278

McCullough, B. D. (1997), A Review of RATS v4.2: Benchmarking Numerical Accuracy, *Journal of Applied Econometrics* 12, 181-190

McCullough, B. D. (1998) Assessing the Reliability of Statistical Software: Part I, *The American Statistician*, 52, 358-366

McCullough, B. D. (1999), Assessing the Reliability of Statistical Software: Part II, *The American Statistician*, 53, 149-159

McCullough, B. D. 1999b. Econometric Software Reliability: EViews, LIMDEP, TSP and SHAZAM, *Journal of Applied Econometrics*, forthcoming.

McCullough, B. D. (2000) Wilkinson's Tests and Econometric Software, *Journal of Economic and Social Measurement*, forthcoming.

McCullough, B. D. and Charles G. Renfro (1999), Benchmarks and Software Standards: A Case Study of GARCH Procedures, *Journal of Economic and Social Measurement*, forthcoming

Sawitzki, G. (1994a), Testing Numerical Reliability of Data Analysis Systems, *Computational Statistics and Data Analysis* 18, 269-286

Sawitzki, G. (1994b), Report on the Numerical Reliability of Data Analysis Systems, *Computational Statistics and Data Analysis (SSN)*, 18, 289-301

Silk, J. (1996), System Estimation: A Comparison of SAS, SHAZAM, and TSP, *Journal of Applied Econometrics*, 11, 437-450

B.D. McCullough, associate editor
Department of Decision Sciences
LeBow College of Business
Drexel University
Philadelphia, PA 19104
USA

bdmccullough@drexel.edu