Verifying the Solution from a Nonlinear Solver: A Case Study: Reply

By B. D. McCullough and H. D. Vinod*

We are pleased to confirm that any doubt our article (McCullough and Vinod, 2003; hereafter "MV03") may have cast on Ron Shachar and Barry Nalebuff (1999; hereafter "SN99") must be removed. We are especially pleased because we thought it quite unfair that other researchers were able to exempt themselves from such detailed scrutiny. It appears that such researchers no longer will have the luxury of reneging on their agreement to honor the replication policy, as this journal now requires authors of accepted empirical papers to provide all programs and data files for posting on the *AER* Web site as a precondition of publication.

The primary aim of our article (MV03) was to provide a four-part methodology for verifying the solution from a nonlinear solver: check the gradient, examine the trace, analyze the Hessian, and profile the likelihood. We adduced copious evidence (MV03, p. 873) that solvers used by economists can produce inaccurate answers, gave examples of different packages giving different answers to the same nonlinear problems (MV03, p. 874), and showed (MV03, pp. 873–74), at least in this journal, that researchers make no effort to verify the solutions from the solvers that they use. We believe this uncritical acceptance of solutions from nonlinear solvers to be a systemic problem in economic research; that is why we wrote the article—certainly, econometrics texts do not show how to verify the solution from a nonlinear solver. In passing, we also showed how a problem can be too large for conventional PC methods, and indicated the failure of replication policies in this journal and other journals.

We used the data and likelihood function from SN99 to illustrate the methodology. In the course of this illustration, we noted that the Hessian was ill-conditioned, suggested that there might exist multiple optima, and that inference based on the Wald statistic was not appropriate. We concluded that the solution we found was, at best, a tentative solution. However, as shown in Shachar and Nalebuff (2004; hereafter SN04), when the problem is rescaled the Hessian is not ill-conditioned; they have correctly identified the difference between the condition number of the badly scaled version of the problem that we analyzed in MV03 and the well-scaled problem that they have analyzed.[1] When the problem is correctly scaled, the Hessian is well-conditioned, the model is locally identifiable, the problem can be solved on a PC, a solution to the problem exists, and SN present it in their Table 1.

Though we were aware that rescaling could ameliorate ill-conditioning (MV03, p. 882), we were unaware of the distinction between artificial ill-conditioning and inherent ill-conditioning, so the method we suggested for analyzing the Hessian contained an error of omission. This error caused us to reach an incorrect conclusion concerning the existence of a solution to the problem. We apologize to Professors Shachar and Nalebuff, and we thank them for their gracious understanding in this regard. Accordingly, we have amended our prescription for analyzing the Hessian—see our nearby exchange with David M. Drukker and Vince Wiggins (2004; hereafter "DW") for complete details. Our

* McCullough: Department of Decision Sciences and Department of Economics, Drexel University, Philadelphia, PA 19104 (e-mail: bdmccullough@drexel.edu); Vinod: Department of Economics, Fordham University, Bronx, NY 10458 (e-mail: vinod@fordham.edu). We thank David Letson and Charles Renfro for useful comments and discussions. As before, we are indebted to Professors Shachar and Nalebuff for their commitment to the *AER* replication policy, and we also thank them for comments.

[1] Back in 1999, we asked SN for their data and code, and they promptly sent us some data and code files. As we noted (MV03, p. 888), "Professor Shachar cooperatively and promptly exchanged numerous e-mails with us as we sought to produce a useable data set and understand his code." Despite our collective better efforts, we somehow wound up with a scaling that they did not intend to give us. The problem that SN intended us to have was correctly scaled and well-conditioned.

analysis also suggested two further difficulties, to which we now turn.

Interpreting our original Figure 1 (reproduced as SN04 Figure 3), we remarked (MV03, p. 884) that the "lack of monotonicity suggests the existence of multiple local optima." In fact there is another optimum (we thank SN for confirming our suspicion). Normally this would be a cause for concern—if another optimum exists, does it have a higher likelihood? As SN showed, though, the likelihood is bimodal but the parameter in question is identified only up to sign, and the second mode is therefore irrelevant. Identifying an econometric model in this fashion is discussed in detail by James Hamilton et al. (2003).

Based on the profile-t plots we argued (MV03, pp. 884–85) that Wald statistics are inappropriate, implicitly questioning SN's inferential results. In response, SN presented the likelihood ratio (LR) intervals for three key parameters, and observed that all three are significantly different from zero. SN's inferences thus are sustained. Their Table 2 shows the differences between the Wald and LR intervals, an issue we shall explore in more detail.

Additionally, SN have made two useful contributions. First, in their Section VI they discussed the methods they used to assess the stability and robustness of their results. Any time a solution has been found, the stability and robustness of the solution should be investigated. Such investigations should be a part of most empirical articles using either linear or nonlinear methods, and we hope they become commonplace. Second, SN provided a textbook example of reparameterizing a nonlinear model to induce a quadratic profile.

The likelihood profiles of some of the parameters are not quadratic, and this leads to a divergence between the Wald and LR intervals; we explore this in Section I. We wish to stress that we never alleged or implied that *any* of SN's results were not replicable; to the contrary, we specifically noted (MV03, footnote 20) that, subject to the usual rounding error and typographical error exceptions,[2] we successfully replicated several of their tables; we discuss this in Section II. Section III presents the conclusions.

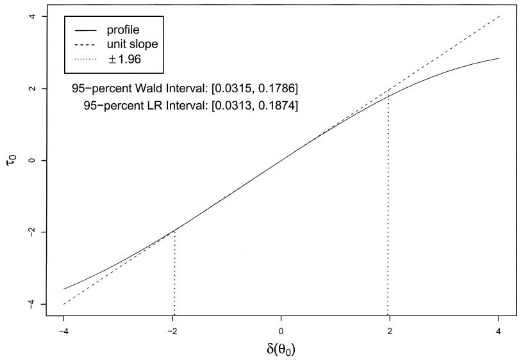

FIGURE 1. PROFILE-t PLOT OF η

I. LR Intervals and Profiling

When, for purposes of this reply, we decided that we wanted to profile and to calculate LR intervals for all 42 parameters, we immediately upgraded our OS/software combination so that we could run GAUSS on this problem. Both profiling and calculation of LR intervals happen to be particularly easy with GAUSS software.

Many of the parameters exhibited quadratic profiles, for example b_0, displayed in Figure 2 below. Not only does the profile-t plot[3] fall almost directly on the line with unit slope, the LR and Wald intervals agree to two digits. Even slight discrepancies from the line with unit slope can induce appreciable discrepancies between the Wald and LR intervals, e.g., the parameter η, given in our Figure 1. For this parameter, the Wald intervals and LR intervals are sufficiently disparate that they would lead to different coverage levels in a Monte Carlo study.

[2] Exact duplication of several digits is to be expected for linear procedures. The same is not true for nonlinear procedures, where only a few digits may be the same, depending on whether different packages, or even different versions of the same package, are used. For example, if one researcher reported −0.8581 for a coefficient of a nonlinear problem and another researcher estimated −0.8589, we would consider that to be close enough and call it "rounding error" (even though it is technically not rounding error). This presumes that a difference in the fourth decimal of one coefficient is immaterial. And, of course, there is the exception for typographical errors.

[3] We remind the reader that the profile can be linearized and standardized so that *if* the original profile *is* quadratic, then the profile-t plot (i.e., the linearized and standardized version) should fall on a line through the origin with unit slope.

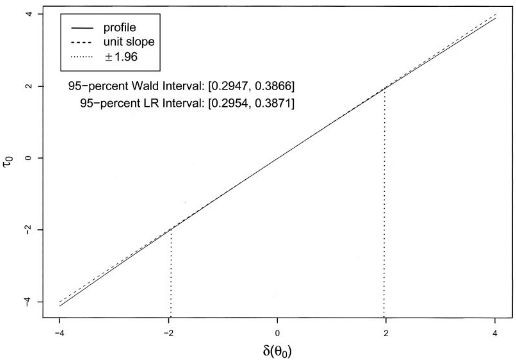

FIGURE 2. PROFILE-t OF b_0

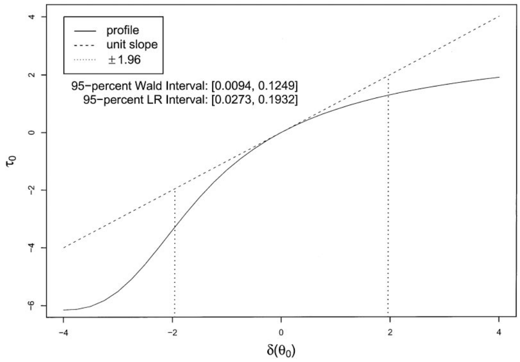

FIGURE 4. PROFILE-t OF S

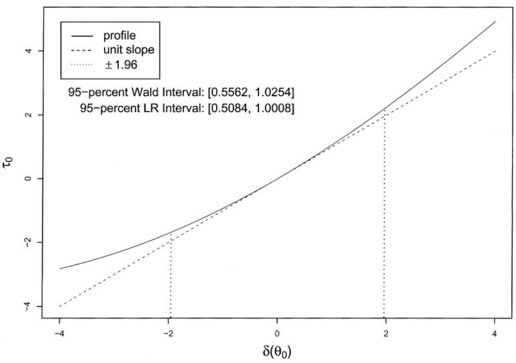

FIGURE 3. PROFILE-t OF ρ

To see this, suppose that the quadratic approximation was exact so that the nominal and actual coverage of the Wald interval was 95 percent. In the case of η, the Wald interval is [0.0315, 0.1786], giving 2.5 percent in each tail. If we were to superimpose the lower and upper limits of the corresponding LR interval[4] [0.0313, 0.1874] onto the Wald interval, to the left of 0.0313 would be a 2.4-percent tail and to the right of 0.1874 would be a 3.8-percent tail. This suggests a difference in the coverage probabilities between the Wald and LR intervals, and the situation is similar for the parameters ρ and S, displayed in our Figures 3 and 4. For ρ, the left and right tails would have 1 percent and 3.9 percent, respectively, and for S, 8.7 percent and 0.0 percent in the left and right tails, respectively. SN did apply an LR test to the parameter S in their original article.

The reparameterization of the model (SN04, Sec. V) has a dramatic effect on the shape of the likelihood surface. A key component of the reparameterization is the transformation $\beta_3 = -\rho\eta$. The profile-t plot of ρ is given in our Figure 3, and that of η in our Figure 1. Both are decidedly not quadratic. Yet, the profile of β_3 in SN04 Figure 5 appears to be quadratic, and the profile-t plot of β_3 in SN04 Figure 6 shows that it is exactly so: the profile-t plot falls precisely on the line with unit slope. This is the essence of reparameterizing to induce quadratic behavior. In addition to improving the quality of inference when one relies on the Wald statistic, such reparameterizations can also enhance the ability of a nonlinear solver to find a solution.

II. Replication and Replication Policies

We thought it rather clear that we could not run GAUSS on SN's nonlinear model due to a log-of-zero problem that was also encountered by the developer of GAUSS (MV03, p. 878), that this was an operating system (OS) problem, and it in no way implied that SN's nonlinear results were not replicable. If we had thought for one second that anyone would conclude, based on our article, that any of SN's results were not replicable, we would have explicitly disclaimed such an errant notion.

SN were correct to insist on an affirmative demonstration of the replicability of their

[4] We temporarily ignore the fact that if the quadratic approximation were exact then the Wald and LR intervals would agree.

results,[5] which demonstration we could not provide. We had reported to SN that their GAUSS code would not run on our computer (we did update both the OS and software one time, to no avail), and so the good Dr. Watson demonstrated what SN knew to be true—that in conjunction with a proper OS/software combination, their code would reproduce their results. For this reply we have again updated our OS/software combination[6] and we can second Dr. Watson's finding that SN's Table 1 is replicable, as well as his comment that SN's replication files for this problem execute "bing-bang-boom." Further, we dug out the files that SN sent us years ago, and we replicated SN99 Table 9, the original table of nonlinear results for their article. Thus, every table in SN99 that we attempted to replicate we successfully replicated.

Susan Feigenbaum and David M. Levy (1993, p. 219) observed that a researcher has two possible defenses against unsuccessful replication: "i) he can work with care so that an article is more likely to be positively replicated; or ii) he can impose costs on potential replicators so that they will do their work more slowly, if at all."

Of the eight empirical articles in the June 1999 issue of this journal, in accordance with this journal's replication policy, four authors provided us with usable data and code (though we only attempted to replicate SN99, who obviously chose the first defense); four did not. The four noncomplying authors, whether by design or by chance, used the second defense: two provided unusable data and code; one admitted he had lost the data; and the fourth pleaded that he did not have time to take the data and code off his hard drive, but informed us that if we would travel to his school then we could take the data and code off his hard drive.

Dewald et al. (1986; hereafter DTA, p. 590) surmised that the failure of authors to supply usable data and code may be that "authors devote most of their effort to the completion of a manuscript and little to the tedious task of compiling, rechecking, and documenting programs and data" and are therefore unwilling to take the time to provide usable data and code. We do not find this reason to be justifiable: Professors Shachar and Nalebuff had to spend a great deal of time recollecting their files and otherwise assisting us with our request; that did not stop them from making the time necessary to honor the policy.

Regardless of the reasons for noncompliance, it is clear that the decision to supply data and code so that results may be independently verified should not be left in the hands of authors. At present, authors are free to ignore replication policies because the journals impose no penalties for failure to comply. Neither will a "data/code" archive alone do the job because, without a proper incentive scheme, authors can choose the second defense against unsuccessful replication: they can submit partial data, partial code, data that does not run with code, code fragments that do not run, etc., as shown by McCullough et al. (2004). A proper incentive scheme will induce authors to choose the first defense, and this requires that the journal publish failed replications. For this purpose, a replication section such as the one recently instituted by the *Journal of Applied Econometrics* is necessary[7] and it must be used: authors need to fear that their errors will be exposed.

Still, there may be insufficient replication efforts. The market for replication is very limited, as journals typically publish replication efforts aimed at their own journals. If an author undertakes the task of replication and the target journal does not accept the manuscript, to where else can he send the fruits of his efforts? Due to such a limited market, many researchers will not even begin a replication attempt. Therefore,

[5] We remind the reader that William G. Dewald et al. (1986) could only replicate 2 of 54 articles. Analyzing the *Journal of Money, Credit, and Banking* archive (which was nominally mandatory but de facto voluntary), McCullough et al. (2004) found that of more than 150 empirical articles, fewer than 15 were replicable.

[6] Knowing the original OS/software combination is an often overlooked component of reproducing results. We know of a couple packages that have made significant changes to their nonlinear capabilities in recent years: running the same data and code on different versions of the same package produces markedly difference results. Of course, there are troubling cases where running the same code on the same computer can produce different results; see Theodore C. Belding (2000) for details.

[7] The journal will provide a one-page refereed summary of the results, with supporting material placed at the journal's Web site; see Drukker and Weihua Guan (2003) for an example.

alternative venues for publication of replication results are necessary. To increase the number of outlets for (and hence the supply of) replication activity, the *Indian Journal of Economics and Business* has recently instituted a replication section like that of the *Journal of Applied Econometrics,* except that it will take replications of articles from *any* journal; see Teresa D. Harrison (2003), McCullough (2003), and McGeary (2003) for examples. Moreover, the *Journal of Economic and Social Measurement* has just started a replication section that will publish articles on replication attempts about articles in *any* journal. Thus, the results of replication attempts, successful or unsuccessful, will be available to researchers who do literature searches.

In equilibrium, when most researchers have learned to make replication files that run "bing-bang-boom" and to satisfy the other requirements of a successful archive [McCullough et al. (2004) make several recommendations], published replication efforts will be limited to cases where the work is replicable but nonetheless incorrect for some reason such as programming errors, software bugs, or incorrect interpretation of computer output. But merely checking to make sure that published results can be produced by some combination of data and code is the least important reason for an archive (Gary King, 1995b, p. 494). The most important reason is to make it easy for one researcher to build on the work of his predecessors (King, 1995a, p. 445). DTA (p. 600) also noted that "authors realize that their research should build on earlier work, but are forced by the unavailability of original data to employ ad hoc tests for the comparability of their data with those used in previous studies." This journal's adoption of a replication policy did not change this situation, but a properly run archive can. Our experience suggests yet another benefit: such archives can provide a rich source for illustrating new ideas with unexpected subtle revelations. For example, our current exchanges with SN and DW have revealed: (i) the subtle distinction between artificial and inherent ill-conditioning, and how to use eigensystem analysis to improve conditioning; (ii) the existence of bimodal likelihoods where one solution can be discarded on economic grounds; and (iii) how reparameterization can induce quadratic behavior.

In McCullough and Vinod (1999, p. 661) we suggested that replication policies are generally ineffectual, and in MV03 we showed this to be true: we demonstrated the failure of the replication policy at this journal, as well as at *International Journal of Economics* and *International Journal of Industrial Organization.* We hope other journals will join the ranks of the journals that have mandatory data/code archives: *Macroeconomic Dynamics; Journal of Money, Credit, and Banking; Federal Reserve Bank of St. Louis Review;* and now *American Economic Review.*

III. Conclusions

In MV03 we used the SN99 model and data to illustrate a four-part method for verifying the solution from a nonlinear solver. In so doing, we raised three issues concerning their model, one of which was due to an error in our prescription for analyzing the Hessian. In our nearby exchange with DW we have discussed our error at length and appropriately amended our prescription for analyzing the Hessian. Professors Shachar and Nalebuff have more than satisfactorily addressed the remaining two issues, and any doubt that our article (MV03) may have cast on theirs (SN99) must be removed.

Micah Altman et al. (2004) wrote, "The furthering of science depends on reproducible results regardless of the specific field. Without the ability to reproduce research, science is reduced to a matter of faith, faith that what is published is indeed the truth." No one needs any faith to believe that the several tables of SN99 that we replicated are reproducible, for we have demonstrated that they are. For the other three complying authors in our small sample, one needs only a modicum of faith. For the noncomplying authors, there is no reason but blind faith to believe that their results are reproducible and, regrettably, this is true of far too many articles in the field of economics. In response to DTA, this journal (and others) adopted a replication policy. In response to MV03, this journal has adopted an archive. We hope other journals again follow suit.

Computational Details: All calculations were performed using GAUSS v5.0 on a Pentium IV running Red Hat Linux v9.0.

REFERENCES

Altman, Micah; Gill, Jeff and McDonald, Michael. "Recommendations for Replication and Accurate Analysis," in Micah Altman, Jeff Gill, and Michael McDonald, eds., *Numerical issues in statistical computing for the social sciences.* New York: J. Wiley and Sons, 2004.

Belding, Theodore C. "Numerical Replication of Computer Simulations: Some Pitfalls and How to Avoid Them." University of Michigan Center for the Study of Complex Systems, Technical Report CSCS-2000-001, 2000.

Dewald, William G.; Thursby, Jerry and Anderson, Richard G. "Replication in Empirical Economics: The *Journal of Money, Credit and Banking* Project." *American Economic Review*, September 1986, *76*(4), pp. 587–603.

Drukker, David and Guan, Weihua. "Replicating the Results in 'On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators'." *Journal of Applied Econometrics*, January–February, 2003, *18*(1), p. 119.

Drukker, David M. and Wiggins, Vince. "Verifying the Solution from a Nonlinear Solver: A Case Study: Comment." *American Economic Review*, March 2004, *94*(1), pp. 397–99.

Feigenbaum, Susan and Levy, David M. "The Market for (Ir) Reproducible Econometrics." *Social Epistemology*, 1993, *7*(3), pp. 215–32.

Hamilton, James; Waggoner, Dan and Zha, Tao. "Normalization in Econometrics." Unpublished manuscript, UCSD Department of Economics, 2003.

Harrison, Teresa D. "Successful Replication of Thornton's (2000) JMCB Article." *Indian Journal of Economics and Business*, December 2003, *2*(2), pp. 285–86.

King, Gary. "Replication, Replication." *Political Science and Politics*, September 1995a, pp. 444–52.

______ **.** "A Revised Proposal, Proposal." *Political Science and Politics*, September 1995b, pp. 494–99.

McCullough, B. D. "Partially Successful Replication of Brunner's (2000) JMCB Article." *Indian Journal of Economics and Business*, December 2003, *2*(2), pp. 289–90.

McCullough, B. D.; McGeary, Kerry Anne and Harrison, Teresa D. "Lessons from the JMCB Archive." *Journal of Money, Credit, and Banking*, 2004 (forthcoming).

McCullough, B. D. and Vinod, H. D. "The Numerical Reliability of Econometric Software." *Journal of Economic Literature*, December 1999, *37*(4), pp. 633–65.

______ **.** "Verifying the Solution from a Nonlinear Solver: A Case Study." *American Economic Review*, June 2003, *93*(3), pp. 873–92.

McGeary, Kerry Anne. "Successful Replication of Wong's (2000) *Journal of Money, Credit, and Banking* Article." *Indian Journal of Economics and Business*, December 2003, *2*(2), pp. 287–88.

Shachar, Ron and Nalebuff, Barry. "Follow the Leader: Theory and Evidence on Political Participation." *American Economic Review*, June 1999, *89*(3), pp. 525–47.

______ **.** "Verifying the Solution from a Nonlinear Solver: A Case Study: Comment." *American Economic Review*, March 2004, *94*(1), pp. 382–90.