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

By B. D. McCULLOUGH AND H. D. VINOD*

David M. Drukker and Vince Wiggins (2004; hereafter "DW") observed that the solutions to badly scaled and well-scaled versions of the model we analyzed in McCullough and Vinod (2003; hereafter "MV") are practically the same.[1] Analyzing this phenomenon, DW revealed a lacuna in the prescription we offered (MV, Sec. III, subsection B) for analyzing the Hessian. DW called for further research on when the condition number is too high, and correctly indicated that there is a need for determining whether a high condition number "indicates a lack of identification that is inherent in the model and the data, and not something that can be easily fixed[.]"[2] We have conducted the requisite research and can amend our prescription appropriately.

We have seen in the literature numerous times the folk theorem that k digits of accuracy are lost in solving a linear system if the data matrix has condition number 10^k. While it is easy to find the folk theorem, it is not so easy to find an important caveat under which the folk theorem might not hold. We were unaware of this caveat when we wrote our article.

Researching this phenomenon, we found the distinction between artificially and inherently ill-conditioned matrices (G. W. Stewart and J. Sun, 1990, pp. 122–24). An ill-conditioned matrix that can be made well-conditioned by rescaling the columns (i.e., variables) is a special case of the class of artificially ill-conditioned matrices. Roughly speaking, these are ill-conditioned matrices for which it is possible to make the ill-conditioning irrelevant, as far as accuracy of the solution is concerned.[3] For present purposes, an inherently ill-conditioned matrix is one that cannot be made well-conditioned by rescaling the variables (when the matrix is the Hessian from a nonlinear solution, this extends to rescaling the parameters, too). The folk theorem applies to inherently ill-conditioned matrices, not artificially ill-conditioned matrices (Stewart, 1996, p. 120).

Artificial ill-conditioning is sometimes innocuous, as in the present case, but it is not necessarily benign. Artificial ill-conditioning can cause a loss of digits in some component(s) of the solution vector. It can also cause problems for optimization routines (Ajit Shenoy et al., 1998).

There is no mechanistic way to determine whether an ill-conditioned matrix is artificially or inherently ill-conditioned (Stewart, 1996, p. 121). The only way to tell is to try to induce well-conditioning and, if this fails, conclude that the matrix is inherently ill-conditioned. Therefore we amend our prescription for analyzing the Hessian to include the following:

* 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 Joe Shinnerl and Hande Benson for useful discussions, and Marc Nerlove, Roni Shachar, and Barry Nalebuff for comments.

[1] Ron Shachar and Barry Nalebuff (2004; hereafter SN04) have made this observation independently.

[2] DW (footnote 1) distinguish between parameters being (theoretically) "asymptotically identified," and being (numerically) "identified" by the data (in conjunction with the numerical characteristics of the estimation method). We think this distinction is important, but we prefer to keep conditioning (a numerical notion) and identification (a theoretical notion) separate; the term "identified" already has too many uses. Perhaps it would be better to use a perfectly clear expression "numerically estimable," reviving the older term "estimable" for referring to the practical numerical aspects of actually obtaining an estimate of the parameters that are "subject-matter identified" or "economically identified." For example, R. C. Bose (1944) showed that even under perfect collinearity, some linear combinations remain numerically "estimable" while others are not; see also Vinod and Aman Ullah (1981, p. 123).

[3] For example, when the bad scaling is confined to the diagonal elements, a certain form of $PLBL'P'$ factorization (a type of Gaussian elimination for symmetric matrices) is numerically stable and as accurate as if the artificial ill-conditioning had been corrected (see Anders Forsgren et al., 1996, for details).

> The practicioner who encounters a large condition number should examine his data for artificial ill-conditioning before concluding that his results are inaccurate (Stewart and Sun, 1990, p. 124).

In MV we did not perform this important step: we simply observed the large condition number and concluded that the answer was bad. We did not pursue rescaling because we did not think that it was necessary to show how to rescale a problem, and we had specifically stated (MV, p. 882) that rescaling could ameliorate the problem. Since the Hessian of the problem that we analyzed was artificially ill-conditioned and innocuously so, we reached an erroneous conclusion concerning the existence of a solution to the Shachar-Nalebuff (1999; hereafter "SN99") model.[4] Our assertion that the badly scaled model exhausts the limits of a PC remains correct, for in that case the difference between a zero gradient and a nonzero gradient still occurs in the sixteenth digit of the log-likelihood (LogL). Consequently, the only way to tell that the badly scaled solution is valid is to do what DW and SN04 did: rescale the problem, show that the rescaled solution is valid, and then note that the two solutions are practically the same.

In the present case, it was easy to rescale because the leading zeroes in the parameter η were a dead giveaway. In the situation where how to scale is not obvious, how might one proceed? DW (footnote 6) provide a useful heuristic that often works in practice. When a more formal approach is necessary, the eigensystem analysis provides the answer. In the standard eigenvalue analysis of multicollinearity, attention focuses on the parameter corresponding to the dominant component of the eigenvector that is associated with the smallest eigenvalue of the covariance matrix. See Vinod and Ullah (1981, Chs. 1 and 5) for details. In the present case, since the Hessian is the inverse of the covariance matrix, we focus on the dominant parameter of the estimated model indicated by the magnitudes appearing in the eigenvector corresponding to the *largest* eigenvalue. Our original eigensystem analysis (MV, Table 3) gave η as the relevant dominant parameter associated with the dominant eigenvalue, though we used two variables that were incorrectly rounded in about the ninth decimal place.[5] The parameter η remains dominant when we use the data that are not so rounded, as we do now.

For the badly scaled problem the condition number in the infinity norm is 6.1E9. VOTPOP is the variable in the likelihood function that is multiplied by η, so we remove the four leading zeroes from η, divide VOTPOP by 10,000, and resolve the problem, of course remembering to rescale the starting value for the parameter η. The condition number drops to 9.7E7. We then perform another eigensystem analysis; this time the relevant dominant parameter is CM1, which is the coefficient on the variable TRIHEAT. After rescaling and resolving, the condition number drops further to 6.2E7. Continuing in this fashion produces Table 1. For row 7, the dominant parameter is σ_ε, for which there is no variable to scale, so we scale the parameter as indicated in SN04, footnote 11.

For rows zero through 9 setting the convergence tolerance ("tol") to 1E-8 produced a "failure to improve" message. Backing off slightly, e.g., from 1E-8 to 1E-7 produced a convergence message. In each of these cases, though the trace exhibits quadratic convergence, the magnitude of the gradient is large enough to give one pause. Only in the last case were we able to

[4] By contrast, SN04 have properly posed the problem so that it is well-conditioned. See our nearby exchange with them for further details.

[5] The LogL presented by SN04, to 12 digits is: 1967.07143370. The LogL for our TSP solution is: 1967.07143359. Drukker (2003, personal communication) informed us that our data files differed from SN99's data files in the sixteenth digit. Investigating further, we discovered two variables that differed more substantially. We received the SN99 data in several files. When we arranged them into a spreadsheet, we must have inadvertently rounded the variables ZECON41 from ten to eight decimal places and LEGIS from ten to seven decimal places. The correct means for ZECON41 and LEGIS are −0.00442869805 and 9.33508092149, respectively. We used data with means −0.00442869830 and 9.33508092126. This underscores the importance of transmitting not just the variables, but summary statistics on the variables to make sure that the data are correctly received and entered in a package. (We should have asked SN99 for the summary statistics to 16 digits to ensure that we had correctly transferred that data.) The effect of these seemingly small changes is profound. For the badly scaled problem, with the slightly incorrect data, we can reduce the magnitude of the gradient, $\|g\|$, to 8.54E-9. With the correct data, the smallest we can make it is 0.025.

TABLE 1—RESCALING TO REDUCE CONDITION BASED ON THE EIGENSYSTEM

Row	Variable rescaled	"tol"	Condition number	$\|g\|$	LogL	Dominant parameter
0	—	1E-7	6.1E9	0.0281	1967.071433704054 *2432*	η
1	VOTPOP	1E-7	9.6E7	0.000308	1967.071433704054 *0158*	Gallup poll (CM1)
2	TRIHEAT	1E-7	6.2E7	0.000174	1967.071433704054 *0158**	ADA and ACA scores (CM6)
3	ADAACA	1E-7	3.8E7	0.000167	1967.071433704053 *3337*	State legislature (CM9)
4	LEGIS	1E-7	8.8E6	0.000131	1967.071433704053 *7885*	Previous (8 years) vote (CM8)
5	DEV2	1E-7	3.4E6	0.000123	1967.071433704053 *7885**	Previous vote (CM7)
6	DEV3	1E-7	1.2E5	0.000122	1967.071433704054 *0158*	Incumbent (CM3)
7	PRESINC	1E-7	1.1E5	0.000122	1967.071433704053 *3337*	σ_ε (SEP)
8	—	1E-7	1.1E5	0.000042	1967.071433704054 *0158*	GNP growth (CM2)
9	JULYEC2	1E-7	7.8E4	0.000041	1967.071433704053 *3337*	—
10	—	1E-8	7.8E4	4.67028E-12	1967.071433704054 *6980*	—

* Not a typo—identical to the number above.

set "tol" smaller than 1E-7 and still obtain a convergence message. In this case the effect on the magnitude of the gradient is dramatic: it is now as small as we would expect to see at a solution. Given also the well-conditioning of the Hessian, we can declare that row 10 identifies a solution to the problem.

As an aside, recall that we rescaled VOTPOP first and TRIHEAT second. Suppose that we had rescaled LEGIS second, instead of TRIHEAT. Then the condition number would have fallen only to 8.1E7, instead of the 6.2E7 that we did obtain. Thus, the eigenvalue analysis does select the variables whose rescaling will have the greatest effect. After rescaling the variable DEV3, the rescaling of PRESINC, σ_ε, and JULYEC2 have no appreciable effect on either the condition or the magnitude of the gradient, but it would have been a mistake to stop at that point. Rescaling simply to minimize the condition number has little theoretical basis to recommend it (Stewart, 1998, p. 247). Rescaling to eliminate artificial ill-conditioning and/or to drive the magnitude of the gradient sufficiently small, on the other hand, obviously has much to recommend it. It is interesting that all this activity in the condition number and the magnitude of the gradient occurs while the LogL varies only in the sixteenth digit, assuming that the sixteenth digit can be trusted—certainly digits 17–20, indicated in italics, cannot be trusted.[6]

This Comment and Reply, as well as the recent Comment and Reply by Justin McCrary (2002) and Steven D. Levitt (2002), demonstrates the value of making replication files available so that the cumulative body of knowledge can be purged of errors. Yet this purging rarely occurs and scientific progress has been impeded, because researchers can simply refuse to honor requests to supply their data and code and, in so doing, hide errors in their work. They have been able to do so because there was no effective penalty that they might incur for reneging on their agreement to honor the journal's replication policy.

We previously suggested [McCullough and Vinod, (1999, p. 661)] that replication "policies" are honored more often in the breach. We verified this in MV03 (Sec. IV). In response, this journal now requires authors to place their data and code in an archive as a condition of publication. We remark that simply creating an archive and allowing authors to place in it whatever files they so desire is not sufficient to ensure the replicability of published results. This has been shown by McCullough et al. (2004), who analyzed the archive of the *Journal of Money, Credit, and Banking* and also made recommendations for improving the performance of the archive. Nonetheless, just as many journals followed this journal when it adopted a replication policy, we hope that many journals

[6] When TSP executes the command PRINT to print the LogL, it makes a copy of the LogL from where it is stored in memory; this number has only 16 digits. Observe that the LogLs are the same for rows 1, 2, 6, and 8, as well as 3, 7, and 9, and also 4 and 5.

again will follow this journal and adopt mandatory data/code archives.

Computational Details: All calculations were performed using TSP v4.5 on a Pentium IV processor running Red Hat Linux v9.0.

REFERENCES

Bose, R. C. "The Fundamental Theorem of Linear Estimation." *Proceedings of 31st Indian Science Congress*, 1944, *4*, pp. 2–3.

Drukker, David M. Personal communication, 2003.

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.

Forsgren, Anders; Gill, Philip E. and Shinnerl, Joseph R. "Stability of Symmetric Ill-Conditioned Systems Arising in Interior Methods for Constrained Optimization." *Siam Journal on Matrix Analysis*, 1996, *17*(1), pp. 187–211.

Levitt, Steven D. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply." *American Economic Review*, September 2002, *92*(4), pp. 1244–50.

McCrary, Justin. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment." *American Economic Review*, September 2002, *92*(4), pp. 1236–43.

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*, June 1999, *37*(2), pp. 663–65.

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

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.

Shenoy, Ajit; Heinkenschloss, Matthias and Cliff, Eugene. "Airfoil Design by an All-At-Once Method." *International Journal for Computational Fluid Mechanics*, 1998, *11*, pp. 3–25.

Stewart, G. W. *Afternotes on numerical analysis.* Philadelphia: SIAM, 1996.

______. *Matrix algorithms, volume 1*. Philadelphia: SIAM, 1998.

Stewart, G. W. and Sun, J. *Matrix perturbation theory*. New York: Academic Press, 1990.

Vinod, H. D. and Ullah, Aman. *Recent advances in regression methods.* New York: Marcel Dekker, 1981.