Journal of Economic Perspectives—Volume 17, Number 1—Winter 2003—Pages 223–227

Comments

To be considered for publication in the "Comments" section, letters should be relatively short—generally fewer than 1000 words—and should be sent to the journal offices at the address appearing inside the front cover. The editors will choose which letters will be published. All published letters will be subject to editing for style and length.

Econometrics and Software

While we enjoyed reading the various articles in the Fall 2001 "Symposium on Econometric Tools," we were chagrined to find that not a single author mentioned that econometric results can be software dependent, a point we recently tried to make (McCullough and Vinod, 1999). Of course, when two packages give different answers to the same problem, it is difficult to know which answer to trust. Obtaining reliable solutions from a nonlinear solver is more than simply coaxing the software package into declaring convergence, another point we have tried to make (McCullough and Vinod, 2003). For example, it is not safe to rely upon default options when conducting nonlinear estimation (McCullough, 1999).

We consider Professor Engle's "GARCH 101" article because it well illustrates both our points; we thank him for providing his data and code. We had no trouble replicating his EViews v3.1 GARCH(1,1) results for the PORT variable given in his Table 3: c = 1.40E-06, $\alpha = 0.0772$, and $\beta = 0.9046$. This result was obtained using default options. If we simply tighten the convergence tolerance from the default 1E-3 to 1E-7, the coefficients change dramatically, as shown in Table 1. Which of these two "solutions" is correct? Is there a third solution that is even more correct than the other two?

Of course, we understand why Professor Engle ran his software at default. He was writing an expository tutorial on GARCH, not on nonlinear estimation. If he had run his package at other than default, he would have had to explain why, and that would have taken him too far afield. Still, we wish that all the authors in the symposium had offered some advice on choosing and using software for the econometric methods they presented.

Returning to the example, not only does changing the tolerance affect the "solution" offered by the nonlinear solver, in the present case switching packages has a similar effect, as shown in our table. Possible reasons for these different estimates include differences in starting values, derivative calculation (analytic versus numeric), maximization routine and other reasons discussed in McCullough and Renfro (2000). Perhaps the most startling reason is that the packages might well be maximizing different likelihood functions, a point that is often glossed over in user guides and textbooks, but that is explored in McCullough and Renfro (1999).

In the context of value-at-risk, different coefficients lead to different residuals and different forecasts of conditional variance and, ultimately, to different estimates of value-at-risk. For ease of comparison, we incorrectly assume that all packages estimate the same forecast standard error as given by Engle, which is 0.0146. For further ease of comparison, we take the first percentile of the standardized residuals from the various packages to be the 27th of the 2608 residuals. Different packages often use different methods to compute quantiles, and some methods are better than others. Hyndman and Fan (1996) present a comparison of six different methods.

Multiplying the first percentile by the forecast standard error and then multiplying by \$1,000,000 gives the value-at-risk, which clearly differs depending upon the package used. The value-at-risk can range from \$39,454 to \$41,033, a difference of \$1,579. Which of these various numbers is preferable? For the alphabet soup of

Table 1
Econometric Software and Varying Estimates

Package	*Tol*	*Iters*	*c*	α_1	β_1	*1st Perc.*	*VaR*
EViews 3.1	1E-7	34	1.09E-06	0.0654	0.9202	2.7023	$39,454
TSP 4.5	1E-7	9	1.11E-06	0.0671	0.9178	2.7118	$39,592
R 1.30	1E-9	23	1.11E-06	0.0670	0.9180	2.8105	$41,033
RATS 5.01	1E-7	17	1.25E-06	0.0715	0.9112	2.7911	$40,750
Stata 7.0	1E-7	85	1.26E-06	0.0715	0.9113	2.7942	$40,795

GARCH extensions (E-GARCH, I-GARCH, T-GARCH and so on), we conjecture that different packages giving different answers to the same problem is more the rule than the exception. We are aware of only one benchmark for GARCH estimation: the Fiorentini-Caizolari-Panattoni (1996) GARCH benchmark proposed in McCullough and Renfro (1999). In other words, when two packages give two different answers to an E-GARCH, I-GARCH or T-GARCH model, we have absolutely no idea which one is correct. Brooks, Burke and Persand (2001) document software dependencies with respect to estimation of E-GARCH models.

It would be an interesting exercise for enterprising graduate students to undertake similar analyses of the other articles in the symposium. Because this journal has a replication policy, the authors of the other articles in the symposium will, as did Professor Engle, make available their data and code. We look forward to reading about the results.

B. D. McCullough
Drexel University
Philadelphia, Pennsylvania

H. D. Vinod
Fordham University
Bronx, New York

References

Brooks, Chris, Simon Burke and Gita Persand. 2001. "Benchmarks and the Accuracy of GARCH Model Estimation." *International Journal of Forecasting.* 17:1, pp. 45–56.

Fiorentini, Gabrielle, Giorgio Calzolari and Lorenzo Panattoni. 1996. "Analytic Derivatives and the Computation of GARCH Estimates." *Journal of Applied Econometrics.* 11:4, pp. 399–417.

Hyndman, Rob and Yanan Fan. 1996. "Sample Quantiles in Statistical Packages." *American Statistician.* 50:4, pp. 361–65.

McCullough, B. D. 1999. "Econometric Software Reliability: E-Views, LIMDEP, SHAZAM, and TSP." *Journal of Applied Econometrics.* 14:2, pp. 191–202.

McCullough, B. D. 2000. "Econometric Software Reliability: E-Views, LIMDEP, SHAZAM, and TSP: Comment and Reply." *Journal of Applied Econometrics.* 15:1, pp. 107–11.

McCullough, B. D. and Charles G. Renfro. 1999. "Benchmarks and Software Standards: A Case Study of GARCH Procedures." *Journal of Economic and Social Measurement.* 25:2, pp. 59–71.

McCullough, B. D. and Charles G. Renfro. 2000. "Some Numerical Aspects of Nonlinear Estimation." *Journal of Economic and Social Measurement.* 26:1, pp. 63–77.

McCullough, B. D. and H. D. Vinod. 1999. "The Numerical Reliability of Econometric Software." *Journal of Economic Literature.* 37:2, pp. 633–65.

McCullough, B. D. and H. D. Vinod. 2003. "Verifying the Solution from a Nonlinear Solver: A Case Study." *American Economic Review.* Forthcoming.

* * *

Instrumental Variables

In their recent survey of instrumental variables and natural experiments, Angrist and Krueger (Fall 2001, pp. 69–86) claim, as has been asserted elsewhere in the literature, that the bias of two-stage least squares is proportional to the degree of over-identification so that bias will increase as the number of (over-identifying) instruments increases (p. 79). The assertion is false because the factor of proportionality is not constant, but can vary with the number of instruments. The misconception probably originated with a misreading of Nagar's (1959) derivation of the approximate bias of the k-class estimator and was then reinforced by tabulations derived from the exact distribution results of Richardson (1968), Sawa (1969) and Phillips (1980). Textbooks have subsequently incorporated these results and thus ensured the dissemination of the error; see, for example, Davidson and MacKinnon (1993, p. 222).

The lack of constancy of the factor of proportionality and the resultant indeterminacy of the direction of instrumental variable bias was first noted in Buse (1992), where I generalized Nagar's (1959) results to subsets of instruments. One of the results of that enquiry, however, showed that bias must increase with the number of instruments if the instruments are irrelevant to the model. Irrelevant and weak instruments live in the same neighborhood, so the assertion by Angrist and Krueger about bias and the number of instruments is likely to be true for the natural experiment literature, where weak instruments are pervasive. But without this qualification, their assertion remains potentially misleading.

To put some flesh on these skeletal remarks, let me offer an heuristic interpretation of my generalization of Nagar (1959) when there is only one endogenous regressor in the structural equation, the usual case in the natural experiment literature. In these circumstances, the explained sum of squares in the reduced form regression of the endogenous regressor on the selected instruments appears in the denominator of the factor of proportionality alluded to earlier. Adding instruments will in general increase the explained sum of squares, and thus both numerator and denominator of the expression for the approximate bias will increase. The change in the direction of the bias is therefore indeterminate. Large increases in the explained sum of squares are likely to reduce bias and conversely for small increases. If the additional instruments are irrelevant, then the explained sum of squares does not change, and bias will increase.

Similar conclusions emerge when the papers on the exact distribution of instrumental variable estimators by Richardson, Sawa and Phillips are re-examined; the mathematical detail for Phillips's results are given in Buse (1992, p. 174). The tabulations in these papers all show increasing bias as the number of instruments increase, but all these tabulations were made with a "concentration" parameter held constant. The concentration parameter can be interpreted heuristically as the explained sum of squares in the reduced form regression of the included endogenous regressors on the chosen instruments. The combined effect of more instruments and a larger explained sum of squares on the bias is again indeterminate. And again, irrelevant instruments will increase bias.

The possibility of decreased bias with additional instruments is more than an analytical curiosum. In the case of one endogenous regressor, the ratio of squared biases for alternative instruments can be determined empirically (Buse, 1992, p. 79). I applied this result to Klein's Model I investment equation to illustrate a reduction in bias with increasing instruments. Another example can be found in the Monte Carlo evaluation of instrumental variable estimators for a single structural equation in a system with autocorrelation by Buse and Moazzami (1991), where in one set of comparisons increasing the number of instruments reduced bias in one-third of the cases.

The message here is not inconsistent with that given by Angrist and Krueger: Pay attention to the fit of the reduced form. In addition, think hard about the order of instrument selection if that luxury is available to you. Adding weak instruments with little explanatory power will likely increase bias. On the other hand, if you start with weak instruments but then have the good fortune to find an instrument(s) with strong explanatory power (in the reduced form), then you should definitely use it, as it is quite likely to reduce bias. For a recent discussion on the statistical evaluation of instrument relevance, see Davis and Kim (2002) and the references cited therein.

I have a concluding observation about the history of weak instruments. Insofar as there is no discussion in Nagar (1959) about the degree of correlation between instruments and endogenous regressors, the reference to Nagar in this context by Angrist and Krueger (p. 74) is unwarranted. As best as I can determine, the first explicit discussion appears in Nelson and Startz (1990).

A. Buse
University of Alberta
Edmonton, Alberta, Canada

References

Buse, A. 1992. "The Bias of Instrumental Variables." *Econometrica.* 60:1, pp. 173–80.

Buse, A. and B. Moazzami. 1991. "Some Evidence on the Potential of Theil's Generalized Two Stage Least Squares Estimator." *Empirical Economics.* 16:3, pp. 335–50.

Davidson, R. and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* Oxford, U.K.: Oxford University Press.

Davis, G. C. and S. Y. Kim. 2002. "Measuring Instrument Relevance in the Simple Endogenous Regressor-Multiple Instrument Case: A Simplifying Procedure." *Economics Letters.* 74:3, pp. 321–25.

Nagar, A.L. 1959. "The Bias and Moment Matrix of the General k-Class Estimator of the Parameters in Simultaneous Equations." *Econometrica.* 27:4, pp. 575–95.

Nelson, C. R. and R. Startz. 1990. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica.* 58:4, pp. 967–76.

Phillips, P. C. B. 1980. "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing n+1 Endogenous Variables." *Econometrica.* 48:4, pp. 861–78.

Richardson, D. H. 1968. "The Exact Distribution of a Structural Coefficient Estimator." *Journal of the American Statistical Association.* 63:4, pp. 1214–216.

Sawa, T. 1969. "The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimators." *Journal of the American Statistical Association.* 64:3, pp. 923–37.

* * *

The Environmental Kuznets Curve

In their assessment of the impact of globalization on the environment ("Confronting the Environmental Kuznets Curve," Winter 2001, pp. 147–68), the authorial team of Susmita Dasgupta, Benoit Laplante, Hua Wang and David Wheeler may be a bit too quick to side with the optimists. The evidence presented for the optimistic case is not as compelling as the article implies.

To take the most obvious example, contrary to what is claimed in the text, the evidence presented in the article on the relationship between direct foreign investment and pollution levels in China, Mexico, Brazil and the United States appears to support a more pessimistic view. The table below compares the annual rate of decline in pollution levels in the years prior to the take off of foreign direct investment with the years after the take off, as presented in Figures 4 and 5 of his paper, based on the underlying numerical data in Wheeler (2001). In Table 1, the break year is chosen from the data shown in the figures and is given in parenthesis. For the U.S. cities included in the table, 1992 is treated as a break point, since U.S. foreign direct investment rose steeply in subsequent years.

In each case, the rate of decline in pollution levels was slower in the period after the rapid upturn in foreign direct investment than in the period immediately prior to the upturn. While the sample size is extremely small, and the measurements are surely imprecise, these data cannot be presented as supporting the case that foreign direct investment is especially good for the environment.

Table 1

Pollution Growth and Foreign Direct Investment

(annual percentage change)

	Pre Take Off of FDI	*Post Take Off of FDI*
China (1991)	−7.6%	−1.5%
Mexico (1993)	−11.1%	−7.9%
Brazil (1992)	−5.3%	−4.6%
Los Angeles	−7.9%	−1.0%
Chicago	−3.4%	−2.5%
Houston	−1.6%	−0.7%
Atlanta	−6.8%	−1.3%
New York	−5.8%	0.0%

Other claims in the article are equally weak. For example, the article presents evidence that the pollution levels of foreign-owned subsidiaries in developing nations are generally lower than the pollution levels of domestically owned facilities. A comparison based on ownership ignores the obvious possibility that multinational corporations may prefer contractual relationships, rather than direct ownership, with the worst polluters, as has been the case with factories that are notorious for especially bad labor conditions (Chan, 1995).

The evidence presented in this article to support the optimistic scenario of declining pollution levels is far less compelling than the discussion implies.

Dean Baker
Helene JorgensenCenter for Economic and Policy Research
Washington, D.C.

References

Chan, A. 1995. "The Emerging Patterns of Industrial Relations in China and the Rise of Two New Labour Movements." *China Information.* 9:4, pp. 36–59.

Wheeler, D. 2001. "Racing to the Bottom? Foreign Investment and Air Pollution in Developing Countries." *Journal of Environment and Development.* 10:3, pp. 225–45.