

# Re-examining 50-year-old OLS estimates of the Klein–Goldberger model

B. D. McCullough\*

*Department of Decision Sciences, LeBow College of Business,  
Drexel University, Philadelphia, PA 19104-2875, USA*

Charles G. Renfro

*Alphametrics Corporation, PO Box 2566, Bala Cynwyd,  
PA 19004-6566, USA*

Houston H. Stokes

*Department of Economics, College of Business Administration,  
University of Illinois at Chicago, 601 S. Morgan Street,  
Chicago, IL 60607, USA*

This paper examines the ordinary least squares estimates of the Klein–Goldberger model by Fox (*Journal of Political Economy*, **64**, 1956, 128). Because Klein and Goldberger published the data set with the model, it is possible to re-examine Fox's results years later, and investigate the accuracy with which these estimates were calculated. The examination reported in this paper was conducted by making independent estimates using three different modern econometric software packages. This examination reveals that the Fox estimates for a number of the equations of this model are replicable, to the two or three digits reported by Fox. Fox's results for other equations cannot be replicated. Not all the reasons for this lack of replicability can be determined, but in several cases the computational methods used by Fox and his assistants have been found to be faulty by modern computational standards.

*Keywords and Phrases:* accuracy, ill-conditioning, least squares, replication.

## 1 Introduction

Economists have been using regression to estimate the values of unknown parameters of econometric models since Tinbergen's first attempts to create his models of the Netherlands and the USA. Regression as a technique can be traced to the work of Gauss and Legendre at the beginning of the 19th century and to its general statistical formalization by Pearson, Yule, and Fisher in the late 19th and early 20th

---

\*bdmccullough@drexel.edu

centuries. In 1939, the idea of using regression techniques for economic applications was sufficiently unusual that Tinbergen devoted the second chapter of volume 1 of his League of Nations study (TINBERGEN, 1939) to a careful description of what is known today by economists as ordinary least squares (OLS).

In the years since, other economists have elaborated on this basic econometric technique, developing variant forms to deal with a variety of particular circumstances. An aspect of this elaboration has been the revolutionary changes that have occurred in the methods of computation used, so that a natural question on the 50th anniversary of the Econometric Institute is, How might these computational changes have affected the results? To shed some light on this issue, in this paper we attempt to replicate OLS parameter estimates by Fox (1956) about the Klein–Goldberger model, which itself also happens to have been created 50 years ago. Fortunately, the data set that underlies this model was published with the model and therefore still exists. Interestingly, it appears that our attempt to replicate Fox’s original estimates is the first replication attempt that has been made. Although the limited information maximum likelihood (LIML) estimates by Klein and Goldberger have been employed in various model performance studies (such as ADELMAN and ADELMAN, 1959; HOWREY, 1971; BOWDEN, 1972), neither the Klein–Goldberger nor the Fox parameter estimation results have been examined during the past 50 years. What we thought would be a straightforward and simple replication proved to be anything but.

## 2 The Klein–Goldberger model

Created at the University of Michigan during the years 1953–1954 and first described in print in 1955 (KLEIN and GOLDBERGER, 1955, hereafter “KG”), the KG model at its inception was pushing the envelope. The electronic computer, nowadays so readily available, was then only just beginning to be used by economists. This model was first estimated using electro-mechanical desk calculators. At that time, neither did computer software yet exist that could be generally applied, nor were stored-program computers accessible: worldwide, only one such machine then existed that was available for general use by economists and that machine was more than 3000 miles away, at the University of Cambridge (RENFRO, 2004a). Undaunted, KG actually made the job more difficult for themselves than absolutely necessary, for they chose to apply LIML, rather than OLS. This choice is quite understandable given the then-recent Cowles Commission focus on simultaneous equations and maximum likelihood as concepts, but nonetheless it implied many more calculations, thereby literally adding days, if not weeks to their work.

On its publication, the KG model was greeted with considerable interest and by the end of 1956 two papers about the book (and also the model) had appeared, one by CHRIST (1956) in the *American Economic Review*, and the other by Fox (1956) in the *Journal of Political Economy*. Of these, that by Fox was somewhat more critical

Table 1. The 15 equations of the Klein–Goldberger model.

Equation
1. Consumer expenditure
2. Gross capital formation
3. Corporate savings
4. Corporate profits
5. Depreciation
6. Private wage bill
7. Private gross national product (GNP)
8. Change in wage rates
9. Value of imports
10. Farm operators' income
11. Farm prices
12. Liquid assets of individuals
13. Liquid assets of businesses
14. Long-term interest rate
15. Short-term interest rate

in its evaluation. Because KG had published not only their model but also all the sample data used, Fox was able to have alternative parameter estimates calculated by OLS, which he then published in the context of his paper. See Table 1 for a list of the 15 stochastic equations, some of which involved relatively complex variable transformations. We focus on Fox’s Table 3 (Fox, 1956, pp. 135–136) which displays, to two and sometimes three decimals, KG’s LIML coefficients and standard errors for two time periods: 1929–1950 (based on the equations given on pp. 50–53 of KG) and 1929–1952 (based on the slightly different equations given on pp. 90–92 of KG), as well as Fox’s OLS coefficients and standard errors for the 1929–1952 period.

### 3 Fox’s OLS

Precisely how Fox obtained his OLS results is unclear. His footnote 7 (Fox, 1956, p. 131) reads, “I am indebted to Mrs. Martha N. Condee, of the Division of Agricultural Economics, U.S. Department of Agriculture, for carrying out most of the least squares computations. Additional computations were supervised by Dr. Emil Jebe, of Iowa State College”. Almost surely these computations were performed with an electro-mechanical desktop calculator. The most popular method for effecting such calculations, Doolittle’s algorithm, was well known to economists of the era. It was featured in KLEIN’S (1953) econometrics text, as well as GOLDBERGER’S (1964) text, although it was known to be of limited accuracy (LEAVENS, 1947). See GOLDBERGER (2004) for a discussion of econometric computing by hand.

Solving for the least squares coefficients of a regression can be carried out in many ways. Doolittle’s method, a variant of Gaussian elimination that lends itself to hand calculation (perhaps with the assistance of a desktop calculating machine), computes what is called an “LU factorization” which is closely related to the Cholesky decomposition. A problem with these methods is that they are not particularly accurate,

especially in the presence of ill-conditioned data, especially when not many digits are carried through the calculations. For ill-conditioned data, the QR decomposition is a much better way to obtain regression estimates.

The effect of the QR decomposition has been illustrated by McCULLOUGH (2000b), who used both the Cholesky and the QR to solve the Wampler1 benchmark regression test (NIST, 1999). Using double precision on a PC (i.e. representing each number to 15 or 16 digits), the QR obtains 10 digits of accuracy, while the Cholesky obtains only six digits of accuracy. The reason the Cholesky is less accurate is that it does not handle cumulated rounding error as well as the QR. Indeed, it is possible for a cumulated rounding error to completely corrupt the solution to a linear regression routine, so that it returns *no* accurate digits. In particular, McCULLOUGH and WILSON (1999) show that this can happen with the linear regression routine in Microsoft Excel.

It is worth recalling the results of LONGLEY's (1967) investigation into the accuracy of early regression routines programmed on computers. After calculating a regression problem by hand, he gave the problem to several existing programs. All but four of the early programs got the answer completely wrong. Of the four "good" programs, one got only the first digit in each of the coefficients, two gave four digit accuracy, and another got the first digit correct for some of the coefficients. Regression programs from the early days of econometrics were not nearly so accurate as the packages we have today.

#### **4 Replication**

As already mentioned, KG provided both the data and the equations for their model, and we set about attempting to reproduce Fox's OLS results using three different software packages: McCullough used RATS v6.2, Renfro used Modler v10.6 Build 22, and Stokes used B34S v8.10Z; features of these and other packages have been described in a recent census of econometric software packages (RENFRO, 2004b). Each of us separately verified the data used and the construction of the equations. However, despite the seeming ease of replicating OLS regressions, we did encounter some difficulties. For example, in his Table 3, Fox presents equation 12 as a standard OLS regression, but we discovered that it was estimated as a log-log model whose intercept was then exponentiated. Further, it was the log-base 10, not the natural logarithm. Nonetheless, we quickly reached agreement with each other. Indeed, we reached the same answers as Fox for equations 3, 4, 9, 11, 13 and 14, shown in Table 2. Curiously, for some of the constant terms but not for others, Fox reported that the standard error is "not available". We can think of no numerical reason that this should occur. We have signified such constant terms by a dagger (†).

Reconciling our results with the rest of Fox's results was not any more easily carried out. For the five equations presented in Table 3, Fox got the slopes correct, but missed the intercept. Our estimates, denoted "MRS", show that we can reproduce

Table 2. Equations Fox got correct.

3. Corporate savings		
Constant		−2.59
Corporate income		0.85
Dividends		−0.24
Corporate savings ( $t - 1$ )		−0.014
4. Corporate profits		
Constant		−8.51
Nonwage farm income		0.72
9. Value of imports		
Constant		0.29
Disposable income		0.0200
Value of imports ( $t - 1$ )		0.54
11. Farm prices		
Constant		−113.36
Price level		2.18
13. Business assets		
Constant†		0.40
Private wage bill		0.22
Interest rate		−0.45
Price change ( $t - [t - 1]$ )		−0.10
Business assets ( $t - 1$ )		0.60
14. Long-term rates		
Constant		2.58
Short term rate ( $t - 3$ )		0.44
Short term rate ( $t - 5$ )		0.26

†signifies “standard error not available”.

Table 3. Equations where Fox missed only the intercept.

Equation/variable	Fox	MRS
5. Depreciation		
Constant	6.47	6.46
Stock of capital	0.09	ok
Private GNP	0.052	ok
7. Private GNP		
Constant†	−22.31	−20.29
Man-years worked	2.02	ok
Stock of capital	0.18	ok
Time trend	2.15	ok
8. Wage rate change		
Constant	6.73	7.22
Unemployment	−1.20	ok
Price change ( $t - 1$ ) ( $t - 2$ )	0.41	ok
Time trend	0.48	ok
12. Individual assets		
Constant†	71.27	71.32
Interest rate	−0.74	ok
15. Short-term rates		
Constant	11.17	11.15
Excess reserves	−0.67	ok

†signifies “standard error not available”.

the slopes but not the intercepts, which suggests that the intercepts were not calculated directly but rather were calculated after the slopes had been calculated. In the three variable case, this would be

$$b_0 = \bar{y} - (b_1 \times \bar{x}_1 + b_2 \times \bar{x}_2).$$

If the slopes were only calculated to a couple decimals, then this inaccuracy could propagate to the intercept. However, this would not explain the observed inaccuracies.

Consider equation 7, given in Table 3, and recall that the correct intercept is  $-20.29$  while Fox reports  $-22.31$ . The mean values of private gross national product (GNP), man-years worked, stock of capital and time trend for the sample period are  $97.006$ ,  $46.6834$ ,  $-2.3775$  and  $10.9$ , respectively. When these mean values are rounded to one digit and combined with the coefficients using the above formula, we get

$$97.0 - (2.02 \times 46.7 + 0.18 \times (-2.4) + 2.15 \times 10.9) = -20.337.$$

When all the numbers are rounded to two significant digits we get

$$97 - (2.0 \times 47 + 0.18 \times (-2.4) + 2.2 \times 11) = -20.768.$$

Rounding to one or two digits produces answers much closer to the truth,  $-20.29$ , than to what Fox reported,  $-22.31$ . Perhaps the formula that Fox used to compute the intercept was in error, or perhaps this is a simple typographical error. Absent specific knowledge of the algorithm used and the number of digits carried, it is all but impossible to reproduce his results precisely.

In two cases, presented in Table 4, Fox not only missed the intercept, but also one of the slope coefficients. When the intercept errors are ignored, which we already have discussed, the slight errors in the slopes are telling. For a problem that is not numerically difficult, such as the regressions in Tables 2–4, a decent algorithm should return several digits of accuracy. Fox's method could not even guarantee three digits of accuracy, so it must have been quite crude by modern standards, which suggests that for more difficult regressions, e.g. regressions with highly collinear data, there may be a tendency for Fox's method to be completely corrupted by rounding error.

The situation gets interesting in the case of the remaining two equations, shown in Table 5, where Fox missed (almost) everything. The discrepancies in equation 10 are sufficiently small that we are comfortable attributing the errors in Fox's results to a mild case of cumulated rounding error. The discrepancies in equation 1 are the cause for concern: it could be a severe case of numerical error, or it could be that

Table 4. Equations where Fox missed the intercept and one slope.

Equation/variable	Fox	MRS
2. Capital formation		
Constant	-16.70	-16.75
Nonwage income	0.78	ok
Stock of capital	-0.073	-0.074
Business liquid assets	0.14	ok
6. Private wage bill		
Constant†	-3.15	-3.06
Private GNP	0.38	ok
Private GNP ( $t-1$ )	0.13	ok
Time trend	0.12	0.11

†signifies "standard error not available".

Table 5. Equations where Fox missed the intercept and slopes.

Equation/variable	Fox	MRS
1. Consumption		
Constant†	-21.06	-59.79
Wage and salary	0.62	0.35
Nonfarm income	0.46	1.04
Farm income	0.39	0.52
Expenditure ( $t - 1$ )	0.21	0.15
Assets ( $t - 1$ )	0.072	0.10
Population	0.25	0.57
10. Farm income		
Constant	0.33	0.36
Nonfarm income	0.010	0.011
Nonfarm income ( $t - 1$ )	0.026	0.023
Agricultural exports	0.013	0.013

†signifies “standard error not available”.

Fox ran one regression but reported another. As a tentative check on the Fox calculations, we re-ran equation 1 using single precision calculation and a Cholesky decomposition. The results agreed to four digits with the results we obtained using our regression packages. This tells us that Fox carried fewer than eight digits through his OLS calculations: if he had carried at least eight, his reported results probably would have been correct to at least two decimals, even using the Doolittle method.

### 5 Further examination

To gain further insight into the inaccuracies in equations 1 and 10, for all 15 equations we computed the summary regression statistics reported by Fox, the  $R^2$  and the standard error of the estimate (s.e.e.). We failed to replicate Fox in only three cases, given in Table 6. Of these, although we were not surprised to be unable to replicate equations 1 and 10 immediately, because of their relative complexity, we were surprised by equation 2, which is on the face of it is a simple specification. None of the discrepancies is large enough to reveal a pattern, so we next computed the condition number for each equation.

The condition number can be a measure of how many digits of accuracy might be lost when computing a regression. When solving a linear system of the form  $Ax = b$ ,

Table 6. Incorrect summary statistics.

Equation	$R^2$	s.e.e.
1. Consumption		
Fox	0.998	1.03
MRS	0.999	0.99
2. Capital formation		
Fox	0.863	3.32
MRS	0.865	3.30
10. Farm income		
Fox	0.269	1.03
MRS	0.265	ok

one significant digit of accuracy might be lost for every power of 10 in the condition number (JUDD, 1998, p. 68). When operating in double precision, a condition number of order  $10E5$  (10 to the fifth power) would present no problem – quite the contrary, however, if operating in single precision. An excellent example is presented by Auckland’s Department of Engineering Science (2003). Consider the system  $Ax = b$

$$\begin{bmatrix} 1.00000 & 0.50000 & 0.33333 & 0.25000 \\ 0.50000 & 0.33333 & 0.25000 & 0.20000 \\ 0.33333 & 0.25000 & 0.20000 & 0.16667 \\ 0.25000 & 0.20000 & 0.16667 & 0.14286 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 2.0 \\ 3.0 \\ 4.0 \end{bmatrix}.$$

The correct solution, obtained using double precision, is given to five digits by  $x' = [-66.514, 927.65, -2585.9, 1862.6]$ . In contrast, single precision provides only three correct digits for each element of  $x$ . If, however, the LU factorization is used and only five digits are carried through the calculations, then the following “solution” is obtained:  $x' = [-85.769, 1157.9, -3160.4, 2245.1]$ . This dramatic change is explained by the condition number of the numeric matrix  $A$  being  $1.6 \times 10E4$ . We see that because its condition number is of order  $10E4$ , if solution is attempted carrying only five digits, the answer will involve *no* accurate digits.

However, the issue here is not just the computed size of the condition number. As explained in detail in McCULLOUGH and VINOD (2004), for the condition number to have implications for the accuracy of the solution, it is important that the matrix in question be “inherently ill-conditioned” and not “artificially ill-conditioned”. If the matrix is only artificially ill-conditioned, then there is no concomitant loss of accuracy in the solution of the system. In the present case, if a simple rescaling of the variables can change the matrix from ill-conditioned to well conditioned, then the matrix is said to be artificially ill-conditioned. As explained by McCULLOUGH and VINOD (2004), the recommended treatment for distinguishing an artificially ill-conditioned matrix from an inherently ill-conditioned matrix involves an exhaustive eigenvalue analysis of the suspect system. When there are several equations to consider, a useful expedient is recommended by DRUKKER and WIGGINS (2004, footnote 6): scale each variable to have a mean whose absolute value is approximately unity. We adopt this expedient, and compute condition number of the covariance matrix in the  $L_2$  norm as the ratio of the largest eigenvalue to the smallest eigenvalue. Results are presented in Table 7.

It is immediately obvious that according to the usual criterion, equations 1 and 10 are the two equations with the highest condition numbers. However, when the condition number of the rescaled systems is evaluated, it becomes apparent that equation 10 was artificially ill-conditioned, rescaling having reduced its condition number from  $1.9E6$  to  $9.1E3$ . Thus, we conclude that only equation 1 is ill-conditioned. If the condition number is of order  $10E6$  and fewer than eight digits were carried through the calculations, then it is quite plausible that the observed discrepancies in equation 1 might be due to cumulated rounding error. Before reaching a determination as to the source of the inaccuracy of Fox’s results for equation 1, we



Table 7. Condition numbers.

Equation	Original	Rescaled
1. <i>Consumption</i>	3.9E8	3.0E6
2. <i>Capital formation</i>	8.1E4	9.9E2
3. <i>Corporate savings</i>	1.3E4	1.6E3
4. <i>Corporate profits</i>	4.3E3	6.4E1
5. <i>Depreciation</i>	2.3E5	2.6E2
6. <i>Private wage bill</i>	5.5E5	1.5E3
7. <i>Private GNP</i>	3.0E5	5.2E3
8. <i>Wage rate change</i>	3.2E3	1.0E2
9. <i>Value of imports</i>	1.7E5	1.1E3
10. <i>Farm income</i>	1.9E6	9.1E3
11. <i>Farm prices</i>	2.2E5	1.6E2
12. <i>Individual assets</i>	1.9E1	7.2E0
13. <i>Business assets</i>	4.5E4	2.6E2
14. <i>Long-term rates</i>	4.6E1	3.2E1
15. <i>Short-term rates</i>	1.5E3	8.0E0

Table 8. Versions of consumption equation.

Variable	MRS: KG specification	Fox’s results	MRS: Theil specification
Constant†	−59.79	−21.06	−32.54
Wage and salary	0.35	0.62	0.69
Nonfarm income	1.04	0.46	0.50
Farm income	0.52	0.39	0.38
Expenditure ( $t - 1$ )	0.15	0.21	0.21
Liquid assets ( $t - 1$ )	0.10	0.063	0.05
Population	0.57	0.25	0.32
$R^2$	0.9986	0.998	0.9984
s.e.e.	0.9900	1.03	1.0480
Condition number	3.9E8	na	2.2E8
Rescaled	3.0E6	na	1.4E6

†signifies “standard error not available”.

need to examine equation 1 more carefully. In particular, we want to consider the possibility that Fox actually did not report results that correspond to the regression that was performed by his assistants.

For equation 1, the consumption equation, Table 8 shows our estimates using the KG specification, Fox’s published results, and our estimates using an alternative specification given by Theil. Equation 1 (as per p. 90 of KG) is given by

$$C_t = b_0 + b_1(W_1 + W_2 - T_W)_t + b_2(P - T_P - S_P) + b_3(A_1 + A_2 - T_A)_t + b_4C_{t-1} + b_5(L_1)_t + b_6(NP)_t,$$

where “wage and salary income” (the variable corresponding to coefficient  $b_2$ ) is defined as  $P - T_P - S_P$ , with  $P$  as “nonwage nonfarm income”,  $T_P$  as “personal and corporate taxes less transfers associated with nonwage nonfarm income” and  $S_P$  as “corporate savings”. THEIL’s (1971, p. 469) discussion of the KG model gives a slightly different definition of wage and salary income:  $P - T_P - S_P - T_C$  where  $T_C$  is the “corporate income taxes. Although this definition used by Theil is at odds with what is given in KG, it nonetheless provides for a result much closer to Fox’s results, as shown in Table 8.

The consequence of this change is that in addition to all the slope coefficients being much closer to Fox's results, the  $R^2$  now rounds to Fox's  $R^2$  and the s.e.e., once off by 0.04 is now off by less than half that amount. The magnitude of the differences between the Fox results and the MRS results for equations 1 and 10 are approximately the same. With the modified specification the equation is still ill-conditioned of order  $10E6$ . As Fox carried fewer than eight digits through his calculations, even if he did use Theil's specification, his results would still not match ours due to cumulated rounding error. We think it quite possible that Fox did not actually estimate the equation he claimed to estimate.

## 6 Conclusions

We have shown by our examination of Fox's (1956) results that econometric calculations made during the pre-computer age cannot necessarily be taken at face value. We found two instances in which the results are simply not replicable, and other instances in which the results have been marred by computational or numerical problems. Originally, Fox put his estimates forward not just as alternative estimates to those of KG, but as potentially better estimates. In his paper he suggested that, "...the reliability of the coefficients expressing relationships between economic variables is of great importance" and noted that "KG laid considerable stress upon the statistical method by which their coefficients were estimated – the 'limited information maximum likelihood method'". Considering the computationally more tractable OLS estimates that Fox provided – which we know now cannot be accepted at face value – what are we to think of the computationally more complex LIML estimates? Therefore, we need to continue to test and improve the software that economists use, and also to increase the opportunity for published work to be replicated and verified.

However, our main aim is not so much to call attention to problems with Fox's estimates as to illustrate by this example – which in fact we originally only chanced upon in the course of other research – the importance of the independent verification of applied results. Inasmuch as Fox made the point in his paper that KG's choice of estimation method is open to criticism, and he countered by providing his own estimates, it is now educational to realize that the differences between the KG and the Fox estimates are not what Fox indicated. Just as important, if not more so, it is unfortunately the case that today there is too little independent verification of more recent published results. What makes this unfortunate is that there is solid evidence from the few verification studies conducted so far that the computing advances of the past 50 years have not eliminated problems of computational inaccuracy and, in addition, that at least a portion of today's published research cannot be replicated, both in the sense that the work is not replicable and that attempts at replication are only seldom possible. On the basis of these few replication studies which have been performed during the past 20 years (DEWALD, THURSBY and ANDERSON, 1986; McCULLOUGH and VINOD, 2003; McCULLOUGH, McGEARY and HARRISON,

2006a,b; etc), we have reason to believe that some applied research is not replicable. However, it is also true that, because in many cases the original data used, and very often the computational methods employed, are hidden from readers of modern published research, attempts at replication and independent verification cannot generally be performed.

We know that existing software packages can be inaccurate, as we and others have shown. It is not uncommon for different packages to give different answers to the same problems, as McCULLOUGH and RENFRO (1999) have shown for generalized autoregressive conditionally heteroscedastic (GARCH) models, BROOKS, BURKE and PERSAND (2003) for multivariate GARCH models, NEWBOLD, AGIAKLOGLOU and MILLER (1994) for autoregressive moving average models, and STOKES (2004) for probit models. In general, it is not safe to assume that software is accurate (McCULLOUGH, 1998, 1999a,b, 2000a).

Replication is an important area of research for three reasons. First, replication is the cornerstone of science: it purges incorrect results from the cumulative body of knowledge. Secondly, there has been precious little replication performed in economics and, consequently, precious little purging. Thirdly, every time a systematic effort has been undertaken to assess the replicability of economic research, the result has been negative. DEWALD, THURSBY and ANDERSON (1986) analyzed 54 papers from a journal and were able to replicate all the results from only two papers. McCULLOUGH *et al.* (2006a) analyzed the same journal 20 years later and were able to replicate 14 of 62 papers, an improvement to be sure, but hardly one that inspires confidence. McCULLOUGH *et al.* (2006b) analyzed another journal and were only able to replicate nine of 117 papers, which suggests that the need for purging the economics literature is not negligible.

What is necessary to change this? Policies that require authors to make their data and code available upon request do not work. McCULLOUGH and VINOD (2003) analyzed such policies at three journals and found that only six of 15 authors would respond to a request, and this was after making multiple attempts for many of the authors. Rather, mandatory data/code archives are the solution; see ANDERSON *et al.* (2005) for an extended discussion. In response to McCULLOUGH and VINOD (2003), the *American Economic Review* adopted such an archive (BERNANKE, 2004). Shortly thereafter, both *Econometrica* and *Review of Economic Studies* did so, too. More recently, *Journal of Political Economy* has joined the club. We can hope that more journals will follow this trend for, if so, we might finally see the end of software and replicability problems that have plagued economics for the past 50 years.

## References

- ADELMAN, I. and F. L. ADELMAN (1959), Dynamic properties of the Klein–Goldberger model, *Econometrica* **27**, 596–625.
- ANDERSON, R. A., W. H. GREENE, B. D. McCULLOUGH and H. D. VINOD (2005), Data/code archives and the future of economic research. *Manuscript, presented at the 2005 AEA Conference in Philadelphia.*

- BERNANKE, B. (2004), Editorial statement, *American Economic Review* **94**, 404.
- BOWDEN, R. J. (1972), More stochastic properties of the Klein–Goldberger model, *Econometrica* **40**, 87–98.
- BROOKS C., S. P. BURKE and G. PERSAND (2003), Multivariate GARCH models: software choice and estimation issues, *Journal of Applied Econometrics* **18**, 725–734.
- CHRIST, C. (1956), Aggregate econometric models, *American Economic Review* **46**, 385–408.
- DEPARTMENT OF ENGINEERING SCIENCE (2003), *Computational techniques course notes*, Department of Engineering Science, The University of Auckland, Auckland, New Zealand.
- DEWALD, W. G., J. THURSBY and R. G. ANDERSON (1986), Replication in empirical economics: the Journal of Money, Credit and Banking project, *American Economic Review* **76**, 587–603.
- DRUKKER, D. M. and V. WIGGINS (2004), Verifying the solution from a nonlinear solver: a case study: comment, *American Economic Review* **94**, 397–399.
- FOX, K. A. (1956), Econometric models of the United States, *Journal of Political Economy* **64**, 128–142.
- GOLDBERGER, A. (1964), *Econometric theory*, John Wiley and Sons, New York.
- GOLDBERGER, A. (2004), Econometric computing by hand, in: C. G. RENFRO (ed.), *Computational econometrics: its impact on the development of quantitative economics*, IOS Press, Amsterdam, The Netherlands.
- HOWREY, E. P. (1971), Stochastic properties of the Klein–Goldberger model, *Econometrica* **39**, 73–87.
- JUDD, K. L. (1998), *Numerical methods in economics*, MIT Press, Cambridge, MA.
- KLEIN, L. R. (1953), *A textbook of econometrics*, Prentice-Hall, Englewood Cliffs, NJ.
- KLEIN, L. R. and A. S. GOLDBERGER (1955), *An econometric model of the United States*, North-Holland, Amsterdam.
- LEAVENS, D. H. (1947), Accuracy in the Doolittle solution, *Econometrica* **15**, 45–50.
- LONGLEY, J. W. (1967), An appraisal of least squares programs for the electronic computer from the viewpoint of the user, *Journal of the American Statistical Association* **62**, 819–841.
- MCCULLOUGH, B. D. (1998) Assessing the reliability of statistical software: part I, *The American Statistician* **52**, 358–366.
- MCCULLOUGH, B. D. (1999a), Assessing the reliability of statistical software: part II, *The American Statistician* **53**, 149–159.
- MCCULLOUGH, B. D. (1999b), Econometric software reliability: E-Views, LIMDEP, SHAZAM, and TSP, *Journal of Applied Econometrics* **14**, 191–202.
- MCCULLOUGH, B. D. (2000a), Is it safe to assume that software is accurate? *International Journal of Forecasting* **16**, 349–357.
- MCCULLOUGH, B. D. (2000b), Experience with the StRD: application and interpretation, *Computing Science and Statistics* **31**, 16–21.
- MCCULLOUGH, B. D. and C. G. RENFRO (1999), Benchmarks and software standards: a case study of GARCH procedures, *Journal of Economic and Social Measurement* **25**, 59–71.
- MCCULLOUGH, B. D. and H. D. VINOD (1999), The numerical reliability of econometric software, *Journal of Economic Literature* **37**, 633–665.
- MCCULLOUGH, B. D. and H. D. VINOD (2003), Verifying the solution from a nonlinear solver: a case study, *American Economic Review* **93**, 873–892.
- MCCULLOUGH, B. D. and H. D. VINOD (2004), Verifying the solution from a nonlinear solver: a case study: reply, *American Economic Review* **94**, 400–403 [reply to Drukker and Wiggins].
- MCCULLOUGH, B. D. and B. WILSON (1999), On the accuracy of statistical procedures in Microsoft EXCEL 97, *Computational Statistics and Data Analysis* **31**, 27–37.
- MCCULLOUGH, B. D., K. A. MCGEARY and T. D. HARRISON (2006a), Lessons from the JMCB archive, *Journal of Money, Credit and Banking* (to appear).
- MCCULLOUGH, B. D., K. A. MCGEARY and T. D. HARRISON (2006b), Lessons from the Fed. St. Louis Review and other archives. *Manuscript, presented at the 2005 AEA Conference in Philadelphia*.
- NEWBOLD, P., C. AGIAKLOGLOU and J. MILLER (1994), Adventures with ARIMA software, *International Journal of Forecasting* **10**, 573–581.

- NIST (1999), StRD: Statistical reference datasets, <http://www.itl.nist.gov/div898/strd/> (accessed 29 January 2006).
- RENFRO, C. G. (2004a), Econometric software: the first fifty years, in: C. G. RENFRO (ed.), *Computational econometrics: its impact on the development of quantitative economics*, IOS Press, The Netherlands.
- RENFRO, C. G. (2004b), A Compendium of existing econometric software packages, *Journal of Economic and Social Measurement* **29** 359–409.
- STOKES, H. H. (2004), On the advantage of using two or more econometric software systems to solve the same problem, in: C. G. RENFRO (ed.), *Computational econometrics: its impact on the development of quantitative economics*, IOS Press, The Netherlands.
- THEIL, H. (1971), *Principles of econometrics*, Wiley, New York.
- TINBERGEN, J. (1939), *Statistical testing of business cycle theories, II, business cycles in the United States of America, 1919–1932*, League of Nations, Geneva.

Received: November 2005. Revised: December 2005.