

GRETTL 1.6.0 AND ITS NUMERICAL ACCURACY - TECHNICAL SUPPLEMENT

A. Talha YALTA and A. Yasemin YALTA*

Department of Economics, Fordham University, New York

2007-07-13

This supplement provides detailed information regarding both the procedure and the results of our testing GRETL using the univariate summary statistics, analysis of variance, linear regression, and nonlinear least squares benchmarks of the NIST Statistical Reference Datasets (StRD), as well as verification of the random number generator and the accuracy of statistical distributions used for calculating critical values. The numbers of accurate digits (NADs) for testing the StRD datasets are also supplied and compared with several commercial packages namely SAS v6.12, SPSS v7.5, S-Plus v4.0, Stata 7, and Gauss for Windows v3.2.37. We did not perform any accuracy tests of the latest versions of these programs and the results that we supply for packages other than GRETL 1.6.0 are those independently verified and published by McCullough (1999a), McCullough and Wilson (2002), and Vinod (2000) previously. Readers are also referred to McCullough (1999b) and Keeling and Pavur (2007), which examine software accuracy and supply the NADS across four and nine software packages at once respectively. Since all these programs perform calculations using 64-bit floating point arithmetic, discrepancies in results are caused by differences in algorithms used to perform various statistical operations.

1 Numerical tests of the NIST StRD benchmarks

For the StRD reference data sets, NIST provides certified values calculated with multiple precision in 500 digits of accuracy, which were later rounded to 15 significant digits (11 for nonlinear least squares). We compare the number of accurate digits (NADs) calculated by GRETL with the certified values from StRD using the log relative errors defined as:

$$LRE = \max(0, -\log_{10} [|q - c| / |c|]) \quad (1)$$

where q is the value computed by GRETL and c is the NIST certified correct value. In the case that c equals zero, NAD is given by the log absolute error: $LAE = -\log_{10} [|q|]$. For problems with multiple parameters, we choose the least accurate of each vector based on the principle that the strength of the weakest link measures the ultimate strength of a chain.

*Correspondence to: Fordham University, Department of Economics; 441 East Fordham Road; Bronx, NY 10458-9993 E-mail: yalta@fordham.edu

1.1 Analysis of variance

The Analysis of Variance test suite of StRD is composed of eleven different datasets, which are grouped into lower, average, or higher levels of difficulty according to their number of constant leading digits. Since GRETL currently does not have an ANOVA package as such, we created a script to estimate a one-way ANOVA problem and compute F statistics and R^2 as any user of GRETL would. The NADs are supplied in Table 1. Poorer performance for higher difficulty datasets is to be expected and observed with other software as a limitation of 64-bit (double precision) computation, as argued in detail by McCullough (1999a).

1.2 Univariate summary statistics

For the univariate summary statistics, we used GRETL’s “mean(.)” and “sd(.)” commands to calculate means and standard deviations. NIST notes that the lag-1 autocorrelation coefficient “may have several definitions.” Consequently, we used the definition employed by NIST, so that GRETL’s calculations are comparable with the certified values. GRETL performs well in the univariate summary statistics tests. The NADs for the mean (λ_μ), standard deviation (λ_s), and the (lag-1) autocorrelation coefficient (λ_ρ) are given in Table 2 for both GRETL and other econometric packages.

1.3 Linear least squares

By default, GRETL uses the Cholesky decomposition (with an option to use the QR decomposition instead) for solving least squares estimations, and all calculations are done with double precision. Ramanathan (2002) (pp. 163-4) discusses the methodological disagreement among statisticians over the calculation of R^2 where the regression does not have an intercept; we find that GRETL does not employ the same definition adopted by NIST. As a result, we used the definition accepted by NIST for the sake of comparison for the NoInt1 and NoInt2 datasets.

With the default options, GRETL’s results were satisfactory except for the Filip data set where the program failed to produce a solution, giving an error message about the near collinearity problem encountered. On the other hand, when the QR algorithm is used, solutions were produced for all of the eleven data sets and with generally higher NADs in comparison to alternative programs. Table 3 provides the results for the calculation of coefficients (λ_β), standard errors (λ_σ), and R^2 (λ_R) for the QR decomposition. Finally, for solving linear regressions, GRETL has an additional “mpols(.)” command, which employs the GNU multiple-precision library for increased accuracy. While this function is not yet fully integrated into the program, we decided to test it as well. It is worth mentioning that, with this option, GRETL produces an impressive performance where all statistics are uniformly correct to at least 12 digits for the StRD linear regression test suit.

1.4 Nonlinear least squares

The nonlinear regression test suite of StRD includes 27 data sets, which are grouped into three levels of difficulty (8 lower, 11 average, and 8 higher). Solutions are obtained using

128-bit precision, and certified values for coefficient estimates (λ_β), standard errors (λ_σ), and residual sums of squares (λ_r) are reported to 11 decimal places for each data set. Also, StRD provides two sets of starting values for initializing the nonlinear algorithms. We used the first set, which is further from the solution in Euclidian distance and therefore considered more difficult. The second set of starting values were used only when the program fails to produce a solution with the first set.

We used the “nls(.)” command in GRETL, which performs nonlinear least squares using a modified version of the Levenberg-Marquandt algorithm. The user has the option to specify the derivatives of the regression function with respect to each of the parameters. Consequently, we ran two full tests: one in which we supplied the analytical derivatives, and one without doing so. With the analytical derivatives supplied, GRETL achieved convergence with the first set of starting values for all but three data sets, and the results, in general, compare favorably to those of commercial packages. On the other hand, without the analytical derivatives, the program failed to produce a solution for the Bennett5 dataset, and the NADs were lower in several cases, which is to be expected. The results for the Windows version are given in Table 4, which compares GRETL’s performance using analytical derivatives with the “preferred combination” of options for other packages. We also found that the GNU/Linux version of the program, on average, can produce slightly higher NADs, but it also failed to converge with the first set of start values for the MGH09 data set using the default settings. This inconsistency is likely to be caused by the small variations in C libraries and compilers for the two operating systems.

2 Randomness tests of random number generation

As the built-in random number generator, GRETL employs the Mersenne Twister (MT) algorithm proposed by Matsumoto and Nishimura (1998). MT has a super astronomical period of $2^{19937} - 1$ and provides a very high 623-dimensional equidistribution up to 32 bit accuracy. We tested the implementation of the algorithm using the DIEHARD program of Marsaglia (1996), which provides a benchmark for testing random number generators. Choosing the seed 1234, we created three million random numbers and found that GRETL easily passes all of the 18 randomness tests, suggesting that the program is suitable for conducting RNG intensive econometrics such as bootstraps and simulations.

3 Numerical tests of probability distributions

GRETL’s “p-value finder”, which can be accessed from the “Tools” menu, computes the area to the right of input values for the Gaussian, student’s t , chi-square, F , gamma, and binomial distributions. Baiocchi and Distaso (2003) find inaccuracies for this function for version 0.997 of the program, which was reported to be fixed quickly in the subsequent 0.998 revision. We tested the accuracy of the latest version by comparing output from GRETL with Knüsel’s (2003) ELV program, which claims to compute exact values for upper and lower quantiles of nine important statistical distributions for probabilities as small as 10^{-100} . We found that GRETL computes correct values for central probabilities, and the reported values for the extreme tails *seem* to be accurate as well.

References

- Baiocchi G, Distaso W. 2003. GRETLM: Econometric Software for the GNU Generation. *Journal of Applied Econometrics*, **18**: 105–110.
- Keeling KB, Pavur RJ. 2007. A Comparative Study of the Reliability of Nine Statistical Software Packages. *Computational Statistics and Data Analysis*, **51**: 3811 – 3831.
- Leo Knüsel. 2003. Computation of statistical distributions - documentation of program ELV. URL <http://www.stat.uni-muenchen.de/~knuesel>, [Online; retrieved June 8, 2006].
- Marsaglia G. 1996. DIEHARD: a battery test of randomness. URL <http://stat.fsu.edu/pub/diehard>, [Online; retrieved June 8, 2006].
- Matsumoto M, Nishimura T. 1998. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**: 3–30.
- McCullough BD. 1999a. Assessing the reliability of statistical software: Part II. *American Statistician*, **53**: 149–159.
- McCullough BD. 1999b. Econometric software reliability: Eviews, LIMDEP, SHAZAM and TSP. *Journal of Applied Econometrics*, **14**: 191–202.
- McCullough BD, Wilson B. 2002. On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics and Data Analysis*, **40**: 713–721.
- Ramanathan R. 2002. *Introductory Econometrics with Applications*. 5th edn., South-Western College Pub., Ohio.
- Vinod HD. 2000. Review of Gauss for Windows, including its numerical accuracy. *Journal of Applied Econometrics*, **15**: 211–220.

TABLE 1. Numerical Tests of Analysis of Variance

Dataset	Diff.	GRETL	SAS	SPSS	S-PLUS	STATA	GAUSS
		1.6.0	v6.12	v7.5	v4.0	v7	v3.2
		λ_F					
SiRstv	l	13	8.3	9.6	13.3	13.1	12.4
SmLs01	l	15	13.3	15	14.5	14.4	14.5
SmLs02	l	15	11.4	15	14.3	13.3	14.1
SmLs03	l	15	11.8	12.7	12.9	14.7	12.7
AtmWtAg	a	11.7	0	ns	9.7	10.4	8.5
SmLs04	a	9.3	0	0	10.4	10.2	8.5
SmLs05	a	9.3	0	0	10.2	10.2	8.3
SmLs06	a	9.3	0	0	10.2	10.2	6.5
SmLs07	h	3.3	0	0	4.6	4.4	2.7
SmLs08	h	3.3	0	0	2.7	4.4	2.2
SmLs09	h	1	0	0	0	4.2	0

TABLE 2. Numerical Tests of Univariate Summary Statistics

Dataset	Diff.	GRETLL 1.6.0		SAS v6.12		SPSS v7.5		S-PLUS v4.0		STATA v7		GAUSS v3.2							
		λ_μ	λ_s	λ_ρ	λ_μ	λ_s	λ_ρ	λ_μ	λ_s	λ_ρ	λ_μ	λ_s	λ_ρ	λ_μ	λ_s	λ_ρ			
PiDigits	1	15	15	15	15	15	14.7	15	0	15	15	6.8	15	15	14.9	15	15	15	
Lottery	1	15	15	15	15	14.9	15	15	3.4	15	15	7.4	15	15	15	15	15	15	
Lew	1	15	15	15	15	14.8	15	13.2	3	15	15	7	15	15	14.8	15	15	14.8	
Mavro	1	15	13.1	13.7	15	13.1	13.8	15	12.1	4.9	15	13.1	7.1	15	13.1	13.7	15	13.1	13.7
Michelso	a	15	13.9	13.4	15	13.8	13.4	15	12.4	3.4	15	13.8	7.3	15	13.8	13.4	15	13.8	13.4
NumAcc1	a	15	15	15	15	ns	ns	15	15	ns	15	15	15	15	15	15	15	15	15
NumAcc2	a	15	15	15	14	14.2	15	15	15	15	14	15	7.1	15	15	15	14	15	15
NumAcc3	a	15	9.5	15	15	9.5	11.9	15	9.5	15	15	9.5	7.1	15	9.5	11.9	15	9.5	11.2
NumAcc4	h	15	8.3	15	14	8.3	10.7	15	8.3	15	14	8.3	7.3	15	8.3	10.7	14	8.3	9

TABLE 3. Numerical Tests of Linear Regression

Dataset	Diff.	GRETLL 1.6.0		SAS v6.12		SPSS v7.5		S-PLUS v4.0		STATA v7		GAUSS v3.2							
		λ_β	λ_σ	λ_R	λ_β	λ_σ	λ_R	λ_β	λ_σ	λ_R	λ_β	λ_σ	λ_R	λ_β	λ_σ	λ_R			
Norris	1	12.5	14.1	15	12.3	11.9	11.6	12.3	10.2	9.9	12.5	14.1	13.8	12.8	13.5	12.2	10.5	15	-
Pontius	1	12.7	13.1	15	11.4	9.2	8.9	12.5	8.9	8.6	12.7	13.2	12.9	11.5	13	11.6	7.9	14.6	-
NoInt1	a	14.7	14.7	15	14.7	15	11.6	14.7	12.5	12.8	14.7	14.4	14	14.7	15	14.7	13.4	0	-
NoInt2	a	15	14.8	15	15	14.9	15	15	14.3	13	15	15	14.9	15	15	15	14.3	1.2	-
Flip	h	7.5	7.7	11	ns	ns	ns	ns	ns	ns	7.1	7	7.8	ns	ns	ns	0	13.4	-
Longley	h	11.8	13	15	8.6	10.3	10.8	12.1	13.3	13.2	13	14.2	14.1	12.1	12.9	8.5	10	11.4	-
Wampler1	h	9.9	10	15	8.3	15	15	6.6	6.6	15	9.8	15	15	6.9	15	6.1	0	11.4	-
Wampler2	h	13.1	15	15	10	15	15	9.7	9.7	15	13.5	15	15	9.7	15	9.4	4.4	11.4	-
Wampler3	h	9.9	14.1	15	7	10.9	10.8	7.4	10.6	10.8	9.2	13.5	15	6.5	10.8	6.1	6.3	11.4	-
Wampler4	h	7.9	14.1	15	7	11.5	14.8	7.4	10.8	14.2	7.5	13.6	15	6.5	10.8	6.1	10.1	11.4	-
Wampler5	h	5.9	14.1	15	7	11.5	15	5.8	10.8	15	5.5	13.5	15	6.4	10.8	6.1	10.5	11.4	-

