THE ACCURACY OF THE DOMESTIC REGRESSION PACKAGE: THE BLS CASE

by

*Gil R. Rodriguez, Jr.**

INTRODUCTION

Possible errors arising from ordinary least-square computer programs had been extensively analyzed by prominent researchers. Results from the studies of Longley [3] and Wrampler [5] revealed, that for some of the widely used regression packages in various computer types, the estimated regression parameters are not even accurate to the first digit. The nature of solutions of ill-conditioned problems were also examined by Beaton, Rubine and Barone [1] who estimated 1000 regression equations based from a set of "perturbed" data[1] (generated through a random number set approximating a uniform distribution). A major highlight of their study was only 2% of the solutions agreed with the unperturbed solution to one or more digits.

However, despite the repeated warnings of such studies,[2] the accuracy of the OLS and other regression routines in our local computer facilities have never been examined empirically. This paper attempts to alleviate such deficiency by analyzing the precision of regression parameters obtained from the United States Bureau of

*Author is Officer-In-Charge of the Economic Research Division, Bureau of Agricultural Economics, Ministry of Agriculture.

[1]The magnitude of the perturbations ranges from ±.5 of the last digit of Longley's data.

[2]Consult Boehm, Menkhaus and Penn [2].

Labor Statistics OLS routine.[3] It should be noted that the latter (which is in single precision) uses the classical Gram-Schmidt orthogonalization process, i.e.:

$$\hat{B} = V^{-1} N'Y$$

where V^{-1} is upper triangular and $N'N = I$

All computer runs were undertaken at the 128 K IBM computer of the Ministry of Agriculture.

BLS Test Criteria

To achieve the previously-mentioned objective, this paper will utilize the approaches suggested by Mullet and Murray [4], Longley [3] and Wrampler [5]. The sample data used in the initial regression runs is the same set used by Mullet and Murray (M-M), i.e.:

Y	X_1	X_2	X_3
8.0159	2.7147	7.3085	6.7742
7.5229	2.7143	6.9713	5.9269
7.8559	2.4046	6.3256	6.2106
8.4554	3.1610	7.3476	6.8024
7.9170	2.4480	7.4678	7.1608
7.4745	2.4599	6.5169	6.1225
8.0501	2.6868	7.4067	6.8669
8.5484	3.0259	7.6996	7.0876
8.4745	2.8800	7.7096	7.0012
7.9899	3.1380	7.0783	6.3026

The M-M method is summarized by the following steps:

"(i) Regress Y, the so-called dependent variable, on the k independent variables $X_1, X_2, \ldots, X_k$ where $k \leq n$, the sample size, and then

[3]The BLS package had been adapted by Gail Lacy and David E. Kunkel to the IBM 370/125 facilities of the Ministry of Agriculture and to the IBM 360/40 computer of the University of the Philippines (Diliman) under the auspices of Project ADAM.

(ii) Regress $Y + \mathcal{L} X_i$ ($\mathcal{L} \neq 0$) on the same set of k independent variables

(iii) Repeat Step (ii) with different values of $\mathcal{L}$ and different X variables, as desired

The following results in terms of (i) and (ii) are true and can be generalized to include (iii):

(1) The calculated intercept and all slope parameters are invariant in (i) and (ii) except for that of X_i which in (ii) is increased by $\mathcal{L}$.

(2) The residual vector is invariant and, consequently, the error (or residual) sum of squares is also invariant."

On the other hand, the recommended test of Longley is:

(a) Regress Y on the k independent variables, $X_1, X_2, \ldots, X_k$

(b) Regress Y on the k transformed independent variables,

$$X_1 + X_2,\ X_1 - X_2, \ldots, X_k + X_{k+1},\ X_k - X_{k+1}$$

As a result of (a) and (b), the following relations must hold:

$$\begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ B_k \\ B_{k+1} \end{bmatrix} = \begin{bmatrix} B_1' + B_2' \\ B_1' - B_2' \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ B_k' + B_{k+1}' \\ B_k' - B_{k+1}' \end{bmatrix}$$

where the B's are the estimated regression parameters.

In the case of Wrampler, he suggested estimating the following equations in a least squares computer routine:

$$Z_1 = 1 + X + X^2 + X^3 + X^4 + X^5, \; X = 0, 1, 2, 3, \ldots\ldots, 20$$

$$Z_2 = 1 + .1X + .01X^2 + .001X^3 + .0001X^4 + .00001X^5$$

The main justifications of Wrampler for the test model were the ill-conditioned nature of the data set and the predominance in the usage of polynomial equations by researchers in the physical and social sciences. Hence, if the relevant routine is satisfactory, then it must yield an $R^2 = 1$; zero sum of residuals; and the true regression coefficients (which is 1 in the case Z_1).

Empirical Results

The regression parameters obtained through the M-M method are given below:

Dependent Variables	a_o	a_1	a_2	a_3
Y	.8973738	.675709	.388903	.362948
$Y - X_1$	.8973738	−.324291	.388903	.362948
$Y - X_2$	.8973738	.675709	−.611097	.362948
$Y - X_3$	.8973738	.675709	.388903	−.637052

As the results indicate, the BLS routine is quite consistent from 6 to 7 digits.[4] The calculated residual vector is:

Sum of Squared Residuals		Dependent Variable
.077557815	:	Y
.077557815	:	$Y - X_1$
.077557815	:	$Y - X_2$
.077557815	:	$Y - X_3$

[4] It is easy to see that $\alpha = 1$ in the test problem. Also, note that the following is true: $a_i - 1 = a_{ii}$, where $i = 1, 2, 3$. A simple way to prove the previous relation is to examine a_i in a single independent variable equation.

The residual estimate is accurate up to the ninth digit.

The results obtained from the Wrampler model runs through the BLS package are given in Table 1.[5] All the regression estimates are sufficiently close to the true values. Also, the adjusted R^2 and sum of squared residuals obtained were equal to one and "almost" zero, respectively. If one compares the estimates in Table 1 with those derived by Boehm, Menkhaus and Penn (Table 2), we will note that the numerical accuracy of the BLS is acceptable. Furthermore, an application of the M-M test to the Wrampler data indicated the invariant nature of the calculated intercept and the relevant parameters.

The results of the Longley test are given in Table 3. The corresponding entry in each cell is the regression parameter. In all cases examined, the BLS has satisfied the Longley test, e.g., .685197 = 682880 + .002317, etc.

Conclusion and Summary

The preceding results indicate that the accuracy of BLS routine "seems" to be satisfactory within the test criteria and data set considered. A future task will be to test the routine in other computer facilities and to analyze the sensitivity of the computed regression estimates through a simulation approach. The alternative of not pursuing "exogenous" tests is to expand the capabilities of the regression package to detect serious computational problems, e.g., printing the eigen values of the $X'X$ matrix. In the case of the BLS, an index of the ill-conditioned problem is provided by the printing of the error vector[6] from the orthogonalization process encountered in solving for the normal equations and for the standard errors.

---

[5] As recommended by Boehm, Menkhaus and Penn, the order of estimating the parameters of the independent variables was varied to detect any serious rounding errors. However, our regression runs for such cases yielded parameters identical to those of Z_1 and Z_2 equations of Table 1.

[6] It should be a null vector in the absence of severe linearity problems among the independent variables.

Table 1.
Results of Wrampler Equations Estimated Through the BLS Routine

Dependent Variable	b_0	b_1	b_2	b_3	b_4	b_5
Z_1	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
$Z_1 - X$	1.000000	–.000001	1.000000	1.000000	1.000000	1.000000
$Z_1 - X^2$	1.000000	.999999	.000000	1.000000	1.000000	1.000000
$Z_1 - X^3$	1.000000	1.000000	1.000000	.000000	1.000000	1.000000
$Z_1 - X^4$	1.000000	.999999	1.000000	1.000000	.000000	1.000000
$Z_1 - X^5$	.999999	1.000000	1.000000	1.000000	1.000000	.000000
Z_2	1.000000	.100000	.010000	.001000	.000100	.000010
$Z_2 - X$	1.000000	–.900000	.010000	.001000	.000100	.000010
$Z_2 - X^2$	1.000000	.100000	–.990000	.001000	.000100	.000010
$Z_2 - X^3$	1.000000	.100000	.010000	–.999000	.000100	.000010
$Z_2 - X^4$	1.000000	.100000	.010000	.001000	–.999900	.000010
$Z_2 - X^5$	.999999	.100000	.010000	.001000	.000100	–.999990

Dependent Variable	R^2	*Sum of Squared Residuals*[1]
Z_1	1.000000	.10224870 D-22
$Z_1 - X$	1.00000	.36055659 D-12
$Z_1 - X^2$	1.00000	.34988148 D-12
$Z_1 - X^3$	1.00000	.15812345 D-12
$Z_1 - X^4$	1.000000	.44185269 D-12
$Z_1 - X^5$	1.000000	.90717572 D-15
Z_2	1.000000	.14319316 D-21
$Z_2 - X$	1.000000	.13877288 D-23
$Z_2 - X^2$	1.000000	.16685631 D-20
$Z_2 - X^3$	1.000000	.78712358 D-17
$Z_2 - X^4$	1.000000	.12068446 D-14
$Z_2 - X^5$	1.000000	.31435412 D-12

[1]The notation D – refers to the movement of the decimal place to the left of the first digit reported. Hence:

.36055659 D-12 = .000000000000036055659

Table 2

Bochm, Menkhaus and Penn Estimates

Program and Machine	B_0	B_1	B_2	B_3	B_4
BMD_2R					
IBM_1	109.68750	a	a	a	1.12573
CDC	1.68442	a	1.31578	.96155	1.00200
IBM_2	109.68750	a	a	a	1.12573
Σ 7	−637.37500	a	a	9.63999	a
BMD_3R					
IBM_1	101009.125	−1792	928	−112.00	5.0
CDC	1.00099	.99999	1.000	1.0000	1.0000
IBM_2	a	−1792	1616	−96	a
TTLS					
CDC	1.0000	.99997	1.0000	1.0000	1.0000
IBM_2	1.0000	1.00000	1.0000	1.0000	1.0000
LSP					
IBM_2	−510.0625	1097.86689	−407.69507	56.11794	−2.08341
$MDVR_1$					
Σ 7	−128.75	307.6770	−121.05426	18.25566	a
$MDVR_2$					
Σ 7	−160.15883	351.15894	−130.46193	18.81564	a

Note: a means machine did not compute the pertinent regression parameter

Table 2 – Continued

Program and Machine	B_5	R^2	*Sum of Residuals*
BMD_2R			
IBM_1	.99627	1.0	0
CDC	.99996	1.0	.2277
IMB_2	.99627	1.0	0
Σ 7	1.02904	1.0	0
BMD_3R			
IBM_1	.81250	.9165	447902.75
CDC	1.0	1.0	38.26454
IBM_2	a	1.0	0
TTLS			
CDC	1.0	1.0	.025
IBM_2	1.0	1.0	.57892
LSP			
IBM_2	−1.06089	1.0	94214
$MDVR_1$			
Σ 7	1.02030	1.0	0
$MDVR_2$			
Σ 7	1.01980	1.0	423.46462

Table 3

Longley's Results from the BLS Routine

Dependent Variable	*Independent Variables* X_1	X_2	X_3	$X_1 + X_2$	$X_1 - X_2$	$X_2 + X_3$	$X_2 - X_3$	$X_1 + X_3$	$X_1 - X_3$
Y	.685197	.680563							
Y				.682880	.002317				
Y		.286619	.392496						
Y						.339558	–.052938		
Y	.637425		.702917						
Y								.670171	–.032746